

# ESTIMATION METHODS FOR DURATION MODELS

by

Brian P. McCall

University of Minnesota

and

John J. McCall

University of California, Los Angeles

and University of California, Santa Barbara

JANUARY 2005

## TABLE OF CONTENTS

1. Introduction
2. Hazard Functions
3. Counting Processes and Martingales
4. Parametric Methods for Continuous-time Data with Covariates
5. The Cox Regression Model
6. Discrete-Time Duration Data
7. Multi-Spell Discrete-Time Models
8. Competing Risks Models
9. Discrete-time Life History Models
10. Specification Diagnostics for Duration Models

## **Abstract**

This paper is a selected overview of econometric methods for duration models and will appear in the forthcoming book The Economics of Search by the authors. The focus of the paper is on martingale methods for continuous time data and general methods for the analysis of discrete-time data including multi-spell models and general life-history models.

## **1. Introduction**

Many empirical tests of search theory employ duration data (see Devine and Kiefer, 1991). For example, the standard model of job search in a stationary environment model implies that unemployment durations have an exponential distribution. In this chapter we develop some statistical tools used to analyze duration data (for more thorough treatment see Lancaster, 1990 or Klein and Moeschberger, 1997). Duration analysis is also referred to as survivor analysis, where the duration of interest is the survival time of a subject (e.g. person or machine). Much of the recent statistical analyses of duration data focus on the hazard function. The hazard function is related to the probability of exiting the initial state within a short interval, conditional on having survived up to the starting time of the interval. In many applications, hazard functions are conditional on a set of covariates. An important feature of the hazard function is that it can be made to depend on covariates that change over time.

In section 2 of this chapter we shall review the basic definition of a hazard function and its relation to the probability density and cumulative distribution function. Section 3 then gives a brief description of counting process theory and martingales. This framework is useful for analyzing duration data including multiple spell duration data (See Andersen and Borgan, 1985, Arjas, 1989, Fleming and Harrington, 1991, and Anderson, Borgan, Gill and Keiding, 1992 for more thorough discussions). Parametric estimation methods for continuous time duration models with covariates are presented in Section 4 while the semi-parametric Cox regression model is discussed in Section 5. Section 6 presents estimation techniques for grouped or discrete-time duration data. In many situations we are interested in studying an individual's movement through several labor market states over time. After extending discrete-time methods to a multi-spell

framework in section 7 and competing risks models in section 8, section 9 presents the general estimation methods for discrete-time life history data. The chapter concludes with a brief discussion of some specification diagnostic methods that can be derived from the counting process approach.

## **2. Hazard Functions**

This section presents a brief overview of hazard functions. Initially we will focus on models without covariates. Later sections of the chapter will then introduce both time-constant and time-varying covariates into the hazard framework.

Let  $T \geq 0$  represent a positive random duration variable, which has some probability distribution in the population;  $t$  denotes a particular value of  $T$ . In survival analysis,  $T$  is the length of time that an individual lives. In many economic applications  $T$  is the duration of an unemployment spell or the duration of job tenure. The cumulative distribution function (c.d.f.) of  $T$  is defined as

$$F(t) = P(T \leq t), \quad t \geq 0.$$

The survivor function is defined as

$$S(t) \equiv 1 - F(t) = P(T > t).$$

Thus,  $S(t)$  represents the probability that an event has not occurred by time  $t$  or that the individual has “survived past”  $t$ . Throughout this section we assume that  $T$  is continuous and denote the probability density function (p.d.f.) of  $T$  by  $f(t) = \frac{dF}{dt}(t)$ . For  $\Delta t > 0$ ,  $P(t < T < t + \Delta t | T > t)$  is the probability of leaving the initial state in the interval  $[t, t + \Delta t)$  given survival until time  $t$ . The

hazard function for T is defined as

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (1)$$

Thus, the hazard function is the instantaneous rate of leaving per unit time (the “escape” rate)

From equation (1) it follows that, for “small”  $\Delta t$ ,  $P(t \leq T < t + \Delta t | T \geq t) \approx \lambda(t)\Delta t$ . The hazard can then be used to approximate a conditional probability in much the same way that the height of the p.d.f. of T can be used to approximate an unconditional probability. We can express the hazard function in terms of the density and c.d.f. very simply. First, write

$$P(t \leq T < t + \Delta t | T \geq t) = \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{1 - F(t)}.$$

When the c.d.f. is differentiable, we can take the limit of the right hand side, divided by  $\Delta t$ , as  $\Delta t$  approaches zero from above:

$$\lambda(t) = \lim_{h \downarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

Because the derivative of S(t) is -f(t), we have

$$\lambda(t) = \frac{d \log S(t)}{dt} \quad (2)$$

and, using  $F(0) = 0$ , we can integrate (2) to get

$$F(t) = 1 - \exp\left[-\int_0^t \lambda(s)ds\right], \quad t \geq 0 \quad (3)$$

Straight forward differentiation of (3) gives the p.d.f. of T in terms of the hazard function:

$$f(t) = \lambda(t) \exp\left[-\int_0^t \lambda(s)ds\right].$$

Therefore, all probabilities can be computed using the hazard function. For example, for all points  $a_1 < a_2$ ,

$$P(T \geq a_2 | T \geq a_1) = \frac{1 - F(a_2)}{1 - F(a_1)} = \exp\left[-\int_{a_1}^{a_2} \lambda(s)ds\right]$$

and

$$P(a_1 \leq T \leq a_2 | T \geq a_1) = 1 - \exp\left[-\int_{a_1}^{a_2} \lambda(s)ds\right].$$

In many empirical applications the shape of the hazard function is of primary interest. In the simplest case, the hazard function is constant:  $\lambda(t) = \lambda$ , for all  $t \geq 0$ . In this case the exit process is memoryless: the probability of exit in the next interval of time does not depend on how much time has been spent in the current state. The standard continuous-time model of stationary job search with a constant offer arrival rate  $\rho$  and wage distribution  $G$  implies a constant re-employment hazard rate

$$\lambda(t) = \lambda = \rho(1 - G(w^r))$$

For a constant hazard function, equation (3) implies that  $F(t) = 1 - \exp(-\lambda t)$  which is the c.d.f. of the exponential distribution.

When the hazard function is not constant we say that it exhibits duration dependence. Assuming that  $\lambda(t)$  is differentiable, the hazard exhibits positive duration dependence at time  $t$  if  $d\lambda(t)/dt > 0$  and negative duration dependence at time  $t$  if  $d\lambda(t)/dt < 0$ . If  $d\lambda(t)/dt > 0$  for all  $t$  we say the process exhibits positive duration dependence. With positive duration dependence, the probability of exiting the initial state increases the longer one is in the initial state.

**Example 1:** Weibull distribution. A popular parametric distribution used in empirical analysis is the Weibull distribution. The random variable  $T$  is said to have a Weibull distribution= if its c.d.f. is given by  $F(t) = 1 - \exp(-\gamma t^\alpha)$

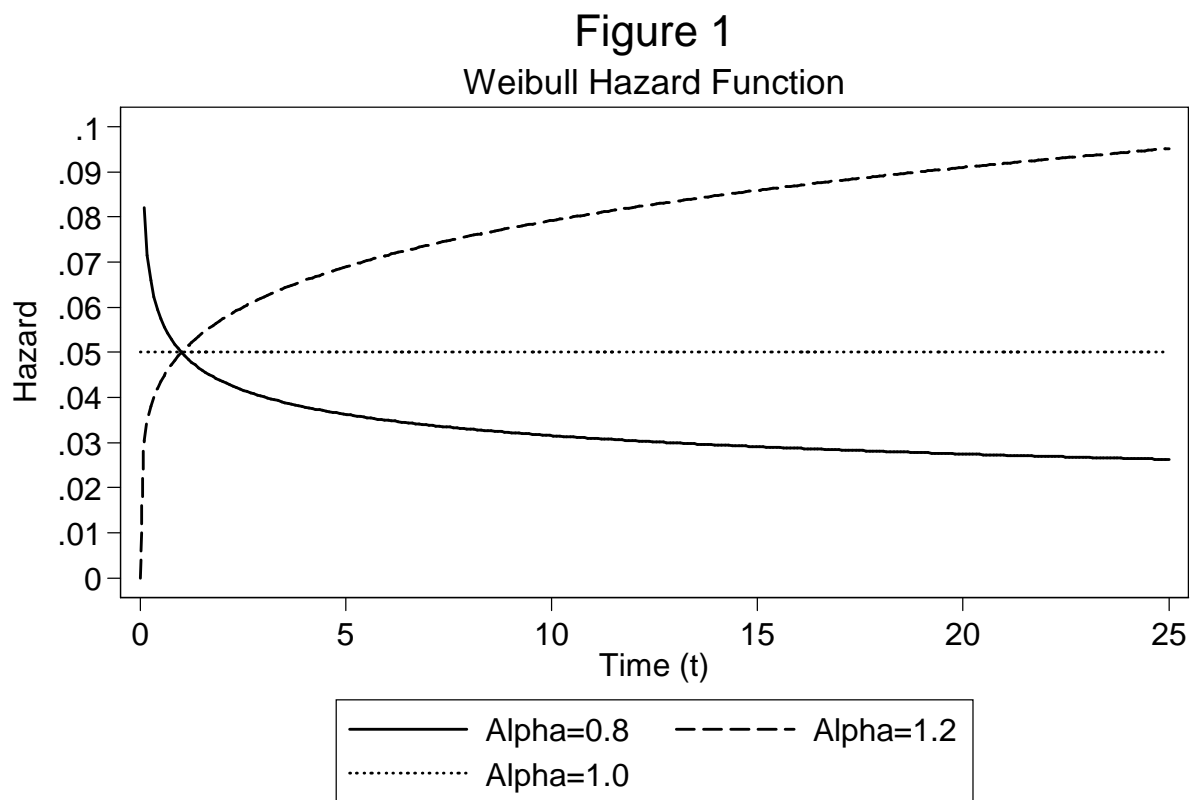
where  $\gamma$  and  $\alpha$  are non-negative parameters. The p.d.f. is given by

$$\lambda(t) = \frac{f(t)}{S(t)} = \gamma \alpha t^{\alpha-1}$$

and the hazard function is  $f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^\alpha)$ . When  $\alpha = 1$ , the Weibull distribution reduces to the exponential with  $\gamma = \lambda$ . If  $\alpha > 1$ , the hazard is monotonically increasing, so the hazard everywhere exhibits positive duration dependence; for  $\alpha < 1$ , the hazard is monotonically decreasing. Graphs of the Weibull hazard function for different values of  $\alpha$  are presented in



Figure1.



Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

**Example 2:** Log-logistic distribution: The random variable T has a log-logistic distribution if its c.d.f. is given by

$$F(t) = 1 - \frac{1}{1 + \phi t^\alpha}$$

and its hazard function is given by

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\alpha \phi t^{\alpha-1}}{1 + \phi t^\alpha} \quad (4)$$

for  $\alpha > 0$  and  $\phi > 0$ . To examine whether the hazard function exhibits positive or negative duration dependence in some ranges we differentiate (4) with respect to t:

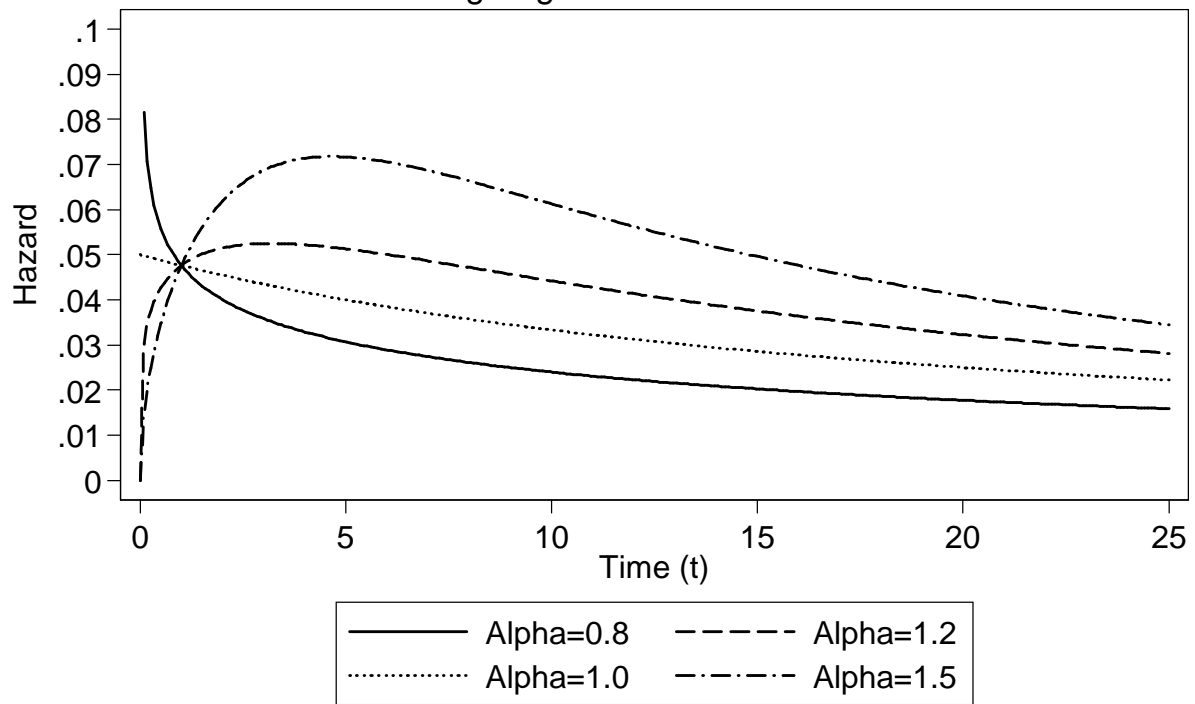
$$\begin{aligned} \lambda'(t) &= \frac{\alpha(\alpha-1)\phi t^{\alpha-2}(1 + \phi t^\alpha) - \alpha^2 \phi^2 t^{2\alpha-2}}{(1 + \phi t^\alpha)^2} = \frac{\alpha^2 \phi t^{\alpha-2}(1 + \phi t^\alpha) - \alpha \phi t^{\alpha-2}(1 + \phi t^\alpha) - \alpha^2 \phi^2 t^{2\alpha-2}}{(1 + \phi t^\alpha)^2} \\ &= \frac{\alpha(\alpha-1)\phi t^{\alpha-2} - \alpha \phi^2 t^{2\alpha-2}}{(1 + \phi t^\alpha)^2} = \frac{\alpha \phi t^{\alpha-2} [(\alpha-1) - \phi t^\alpha]}{(1 + \phi t^\alpha)^2} \end{aligned}$$

For  $\alpha \leq 1$ ,  $\lambda'(t) < 0$  for all t and for  $\alpha > 1$   $\lambda'(t) > 0$  for  $t < \left(\frac{\alpha-1}{\phi}\right)^{1/\alpha}$  and  $\lambda'(t) < 0$  if

$t > \left(\frac{\alpha-1}{\phi}\right)^{1/\alpha}$ . Graphs of the log-logistic hazard function for different values of  $\alpha$  are presented

in Figure 2.

Figure 2  
Log-Logistic Hazard Function



Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

### 3. Counting Processes and Martingales

The theory of counting processes and their accompanying martingales is useful in developing estimation procedures for duration data. Here we will be content with giving a cursory overview of the counting process approach. More detailed discussions can be found in Bremaud (1981), Fleming and Harrington (1991), and Anderson et al (1992).

Recall that a counting process is a process that counts the number of events that occur at random times. Let  $T_n$ ,  $n=0,1,2, \dots$  be a sequence of positive random variables such that

- 1)  $T_0 = 0$
- 2)  $T_n < T_{n+1}$  P a.s.
- 3)  $\lim_{n \rightarrow \infty} T_n = \infty$  P a.s.

Condition 3) is needed to insure that the counting process doesn't blow up to infinity in finite time. The counting process  $N(t)$  is then defined by

$$N(t) = \sum_{n=1}^{\infty} I\{T_n \leq t\}.$$

Thus,  $N(0) = 0$ ,  $N(t) < \infty$  almost surely, and the sample paths of  $N(t)$  are piecewise constant, right continuous and non-decreasing with jumps of size 1.

A filtration or history denoted by  $\{F_t, t \geq 0\}$  is a sequence of sigma algebras indexed by  $t$  that measure the accumulated information up to time  $t$ . As time progresses information increases and so  $F_s \subset F_t$  for  $s < t$ . That is  $A \in F_s \rightarrow A \in F_t$ . The history “just before” time  $t$  is denoted by  $F_{t-}$  and is the sigma algebra generated by all sets in  $F_s$  for all  $s < t$ .

**Definition:** A process  $X(t)$  is adapted to  $\{F_t, t \geq 0\}$  if  $X(t)$  is  $F_t$  measurable for all  $t$ .

Before continuing we shall review some results from martingale theory.

**Definition:** A right-continuous stochastic process  $X(t)$  with left-hand limits is said to be a **martingale** with respect to the history  $F_t$  if

- i)  $X(t)$  is adapted to  $F_t$
- ii)  $X(t)$  is integrable ( $E(|X(t)|) < \infty$ ) for all  $t$
- iii) For all  $0 \leq s \leq t$ ,  $E(X(t)|F_s) = X(s)$  P-a.s.

$X(t)$  is called a **submartingale** if we replace iii) by

- iiia) For all  $0 \leq s \leq t$ ,  $E(X(t)|F_s) \geq X(s)$  P-a.s.

and  $X(t)$  is called a **supermartingale** if we replace iii) by

- iiib) For all  $0 \leq s \leq t$ ,  $E(X(t)|F_s) \leq X(s)$  P-a.s.

The next theorem is important for deriving martingales associated with stochastic processes.

**Doob-Meyer Decomposition Theorem** Let  $X(t)$  be a right-continuous non-negative submartingale with respect to history  $F_t$ . Then there exists a right-continuous martingale  $M(t)$  and an increasing right-continuous predictable process  $A(t)$  such that  $E(A(t)) < \infty$  and

$$X(t) = M(t) + A(t) \text{ P a.s.}$$

**Corollary:** Let  $N(t)$  be a counting process with associated “intensity process”  $\lambda(t)$ . Then

$$M(t) = N(t) - \int_0^t \lambda(s) ds$$

is a  $F_t$  – martingale.

**Definition:** A process  $X(t)$  is said to be predictable with respect to  $\{F_t, t \geq 0\}$  if  $X(t)$  is  $F_{t^-}$  - measurable for all  $t$ .

Another useful theorem is the following:

**Theorem:** If  $V(t)$  is a predictable process such that

$$E\left[\int_0^t |V(s)| \lambda(s) ds < \infty \text{ P a.s. then}\right.$$

$$\int_0^t V(s) dM(s) = \int_0^t V(s) dN(s) - \int_0^t V(s) \lambda(s) ds \quad (5)$$

is a  $F_t$  - martingale.

Let  $C$  be a censoring time and define  $Y(t)$  to be the stochastic process  $Y(t) = I\{C \geq t\}$ .

Thus,  $Y(t)$  equals one up until and including the time at which an observation is censored and equals zero, thereafter. We assume that this stochastic process is measurable with respect to  $F_{t^-}$ .

Further define  $Z(t)$  to be the stochastic process  $Z(t) = I\{T_1 > t\}$ . Using this theorem and defining with  $V(t) = Z(t)Y(t)$  it can then be shown that

$$\int_0^t Y(s)Z(s) dN(s) - \int_0^t Y(s)Z(s) \lambda(s) dt \text{ is a } F_t \text{ - martingale. Thus,}$$

$$E\left(\int_t^{t+\Delta t} Y(s)Z(s) dN(s) \mid F_{t^-}\right) = E\left(\int_t^{t+\Delta t} Y(s) \lambda(s) ds \mid F_{t^-}\right) \quad (6)$$

for all  $\Delta t > 0$ .

This result enables us to derive an estimator for the integrated hazard function  $\Lambda(t)$  :

$$\Lambda(t) \equiv \int_0^t \lambda(s) ds.$$

From (6) we have that

$$E\left[Y(t)(Z(t)dN(t) \mid F_{t^-}\right] \approx Y(t)\lambda(t)E\left[Z(t) \mid F_{t^-}\right]dt. \quad (7)$$

Suppose we have a random sample of size  $N$  and let  $Y_i(t)$ ,  $Z_i(t)$ , and  $N_i(t)$  denote the sample paths of  $Y(t)$ ,  $Z(t)$ , and  $N(t)$  for the  $i$ th individual,  $i = 1, \dots, N$ . Then appealing to the law of large numbers gives:

$$\frac{\sum_{i=1}^N Y_i(t)Z_i(t)dN_i(t)}{N} \approx \frac{\lambda(t)dt \sum_{i=1}^N Y_i(t)Z_i(t)}{N}. \text{ Hence,}$$

$$\lambda(t)dt = \frac{\sum_{i=1}^N Y_i(t)Z_i(t)dN_i(t)}{\sum_{i=1}^N Y_i(t)Z_i(t)} = \frac{\sum_{i=1}^N Y_i(t)Z_i(t)dN_i(t)}{R(t)}$$

where  $R(t)$  is the number “at risk” set at time  $t$  and includes all those who have not been censored or have failed by  $t$ :

$$R(t) = \sum_{i=1}^N Y_i(t)Z_i(t).$$

Thus, we have

$$\hat{\Lambda}(t) = \int_0^t \hat{\lambda}(s)ds = \int_0^t \sum_{i=1}^N \frac{Y_i(s)Z_i(s)}{R(s)} dN_i(s) = \int_0^t \frac{J(s)}{R(s)} dN^*(s) \quad (8)$$

where  $dN^*(s) \equiv \sum_{i=1}^N Z_i(s)Y_i(s)dN_i(s)$  and  $J(s) = 1$  if  $R(s) > 0$  and  $J(s) = 0$  if  $R(s) = 0$  with the

convention that  $0/0=1$ . The estimator  $\hat{\Lambda}(t)$  in (8) is referred to as the Nelson-Aalen estimator of the integrated hazard function and is a step function that is constant at all times except failure times and jumps up by  $1/R(t)$  at time  $t$  when a failure occurs at time  $t$ .

The integrated hazard function of the single duration variable  $T_1$  equals

$\Lambda^*(t) = \int_0^t \lambda^*(s)ds = \int_0^t Z(s)\lambda(s)ds$ . Note this integrated hazard function is stochastic because of  $Z(t)$ .

Thus,

$$\begin{aligned} & \int_0^t \sum_{i=1}^N J(s)Y_i(s)Z_i(s)dN_i(s) - \int_0^t \sum_{i=1}^N J(s)Y_i(s)Z_i(s)d\Lambda(s) \\ &= \int_0^t \sum_{i=1}^N J(s)dN_i^*(s) - \int_0^t \sum_{i=1}^N J(s)Y_i(s)d\Lambda_i^*(s) \\ &= \int_0^t \sum_{i=1}^N J(s)Y_i(s)dM_i^*(s) \end{aligned} \quad (9)$$

So,

$$\hat{\Lambda}(t) - \sum_{i=1}^N \int_0^t \frac{J(s)Y_i(s)}{R(s)} d\Lambda_i^*(s) = \int_0^t \sum_{i=1}^N \left[ \frac{Y_i(s)J(s)Z_i(s)}{R(s)} \right] d(N_i(s) - \Lambda(s)) = \int_0^t \sum_{i=1}^N \frac{Y_i(s)J(s)}{R(s)} dM_i^*(s) \quad (10)$$

Which shows that for all  $t$ ,  $\hat{\Lambda}(t)$  is an unbiased estimator of  $\int_0^t \pi(s)d\Lambda(s)$  where  $\pi(s) = P(J(s) =$

1), since



$$\begin{aligned} \sum_{i=1}^N \int_0^t \frac{J(s)Y_i(s)}{R(s)} d\Lambda_i^*(s) &= \int_0^t \left[ \frac{J(s)}{R(s)} \right] \left[ \sum_{i=1}^N Y_i(s)Z_i(s) \right] d\Lambda(s) = \int_0^t \left[ \frac{J(s)}{R(s)} \right] [R(s)] d\Lambda(s) \\ &= \int_0^t J(s) d\Lambda(s) \end{aligned}$$

Furthermore, as  $N \rightarrow \infty$ ,  $P(J(s) = 1) \rightarrow 1$  a.s. and, hence,  $\hat{\Lambda}(t)$  is a consistent estimator of  $\Lambda(t)$ .

The next two corollaries are useful applications of the Doob-Meyer Decomposition Theorem:

**Corollary 1:** Let  $M(t)$  be a right-continuous martingale with respect to  $F_t$  and assume that

$E[M^2(t)] < \infty$  for all  $t$ . Then there exists a unique right continuous predictable process called the predictable quadratic variation of  $M(t)$  and denoted by  $\langle M, M \rangle(t)$  such that  $\langle M, M \rangle(0) = 0$ ,  $E\langle M, M \rangle(t) < \infty$  for all  $t$  and  $M^2(t) - \langle M, M \rangle(t)$  is a right-continuous martingale.

**Corollary 2:** Let  $M_i(t)$  be a right-continuous martingales with respect to  $F_t$  and assume that

$E[M_i^2(t)] < \infty$  for all  $t$ ,  $i=1,2$ . Then there exists a unique right continuous predictable process called the predictable covariation process of  $M_1(t)$  and  $M_2(t)$  and denoted by  $\langle M_1, M_2 \rangle(t)$  such that  $\langle M_1, M_2 \rangle(0) = 0$ ,  $E\langle M_1, M_2 \rangle(t) < \infty$  for all  $t$  and  $M_1(t)M_2(t) - \langle M_1, M_2 \rangle(t)$  is a right-

continuous martingale and.  $\langle M_1, M_2 \rangle(t)$  is the difference of two increasing right-continuous predictable processes.

Finally, we have

**Theorem:** Let  $V_1(t)$  and  $V_2(t)$  be bounded predictable processes and  $M_1(t)$  and  $M_2(t)$  martingales with respect to  $F_t$  such that  $M_i^2(t) < \infty$ ,  $i=1,2$ . Then

$$\int V_1(t) dM_1(t) \int V_2(t) dM_2(t) - \int V_1(t) V_2(t) d\langle M_1, M_2 \rangle(t) \quad (11)$$

is a martingale. Finally we present a theorem that relates the compensator of  $M^2(t)$  to the compensator of  $M(t)$ .

**Theorem:** Let  $N(t)$  be a counting process with compensator  $A(t)$ . Assume that almost all sample paths of  $A(t)$  are continuous and that  $E[M^2(t)] < \infty$ . Then  $\langle M, M \rangle(t) = A(t)$ . Or in other words,  $M^2(t) - A$  is a martingale.

Sketch of Proof: Integration by parts shows that

$$M^2(t) = 2 \int_0^t M(s^-) dM(s) + \sum_{s \leq t} \{\Delta M(s)\}^2. \text{ Now, since } M(t) = N(t) - A(t) \text{ we have}$$

$\Delta M(s) = \Delta N(s) - \Delta A(s)$ . Substituting into above yields

$$\begin{aligned} M^2(t) &= 2 \int_0^t M(s^-) dM(s) + \sum_{s \leq t} \{\Delta N(s) - \Delta A(s)\}^2 \\ &= 2 \int_0^t M(s^-) dM(s) + \sum_{s \leq t} \{\Delta N(s)\}^2 = 2 \int_0^t M(s^-) dM(s) + N(t) \end{aligned}$$

where the second equality follows from the assumption that  $A(t)$  has no jumps P-a.s. and the third equality follows from the fact that  $N(t)$  is a counting process and so

$$N(t) = \sum_{s \leq t} \Delta N(s) .$$

Since  $M(t) = N(t) - A(t)$  we then have

$$M^2(t) - A(t) = 2 \int_0^t M(s^-) dM(s) + M(t) .$$

Now  $M(s^-)$  is a predictable process, so both terms on the right hand side are martingales which shows that  $A(t)$  is the compensator of  $M^2(t)$ . Q.E.D.

Next, we have

**Theorem:** If  $N_i(t)$ ,  $i=1, \dots, N$  are independent counting processes with compensators,  $A_i(t)$ ,

defined by  $\int_0^t \lambda_i(s) ds$  and  $H_i(t)$  are  $F_t$  – predictable functions then

$M(t) \equiv \sum_{i=1}^N \int_0^t H_i(s) d(N_i(t) - A_i(t))$  is an  $F_t$ -martingale with

1)  $E(M(t)) = 0$  for all  $t$

2)  $\text{var}(M(t)) = \sum_{i=1}^N \int_0^t E[H_i^2(s)] \lambda_i(s) ds$ .

Applying this theorem to the Nelson-Aalen estimator we have

$$\begin{aligned} \text{Var}\left(\hat{\Lambda}(t) - \int_0^t J(s) d\Lambda^*(s)\right) &= \text{Var}\left(\int_0^t \frac{J(s)}{R(s)} dM^*(s)\right) = \\ &= E\left[\int_0^t \left(\frac{J(s)}{R(s)}\right)^2 ds \mid M^*, M^* >\right] = E\left[\int_0^t \left(\frac{J(s)}{R(s)}\right)^2 R(s) d\Lambda(s)\right] \\ &= E\left[\int_0^t \left(\frac{J(s)}{R(s)}\right)(s) d\Lambda(s)\right] \end{aligned}$$

this can be estimated by

$$\int_0^t \left(\frac{J(s)}{R(s)^2}\right) dN^*(s).$$

For large  $n$ ,

$$\lim_{n \rightarrow \infty} n E\left(\hat{\Lambda}(t) - \int_0^t J(s) d\Lambda^*(s)\right)^2 = \lim_{n \rightarrow \infty} E\left[\int_0^t \left(\frac{nJ(s)}{R(s)}\right) d\Lambda^*(s)\right] = \int_0^t \left(\frac{1}{\pi(s)}\right) d\Lambda^*(s)$$

Finally, note that

$$\begin{aligned} \sqrt{n} \left( \hat{\Lambda}(t) - \int_0^t J(s) d\Lambda^*(s) \right) &= \sqrt{n} \left( \int_0^t \frac{J(s)}{R(s)} dM^*(s) \right) = \\ &= \frac{1}{\sqrt{n}} \left( \int_0^t \frac{nJ(s)}{R(s)} d \sum_{i=1}^n M_i^*(s) \right) = \frac{1}{\sqrt{n}} \left( \sum_{i=1}^n \int_0^t \frac{nJ(s)}{R(s)} dM_i^*(s) \right) \end{aligned}$$

which from the martingale central limit theorem leads to the result that the Nelson-Aalen estimator is asymptotically normally distributed.

Summarizing our results for the Nelson-Aalen estimator  $\hat{\Lambda}(t) = \int_0^t \frac{J(s)}{R(s)} dN^*(s)$  we have:

**Theorem:**

- 1)  $\hat{\Lambda}(t)$  is an unbiased estimator of  $\int_0^t J(s) d\Lambda(s)$ .
- 2)  $\hat{\Lambda}(t)$  is a consistent estimator of  $\Lambda(t)$ .
- 3)  $\sqrt{n} \left( \hat{\Lambda}(t) - \Lambda^*(t) \right)$  is asymptotically normally distributed with mean 0 and variance

$$\int_0^t \left( \frac{1}{\pi(s)} \right) d\Lambda^*(s).$$

**i) Kaplan Meier Estimator of the Survivor Function**

The Nelson Aalen estimator can be used to derive an estimator for the survivor function  $S(t)$ .

Note that since

$$d\Lambda^*(t) = \frac{dF(t)}{1-F(t^-)}$$

we have

$$S(t) = 1 - \int_0^t (1 - F(s^-)) d\Lambda^*(s) = 1 - \int_0^t S(s^-) d\Lambda^*(s)$$

So, we can think of an estimator of  $S(t)$  has being defined recursively using

$$\hat{S}(t) = 1 - \int_0^t \hat{S}(s^-) d\hat{\Lambda}(s) \quad (12)$$

where  $\hat{\Lambda}(t)$  is the Nelson-Aalen estimator of the integrated hazard function. Substituting the definition of the Nelson-Aalen estimator into (12) yields:

$$d\hat{S}(t) = \hat{S}(t^-) \left( \frac{J(t) dN^*(t)}{R(t)} \right) = \begin{cases} \hat{S}(t^-) \left( \frac{1}{R(t)} \right) & \text{if } dN^*(t) = 1 \\ 0 & \text{if } dN^*(t) = 0 \end{cases}$$

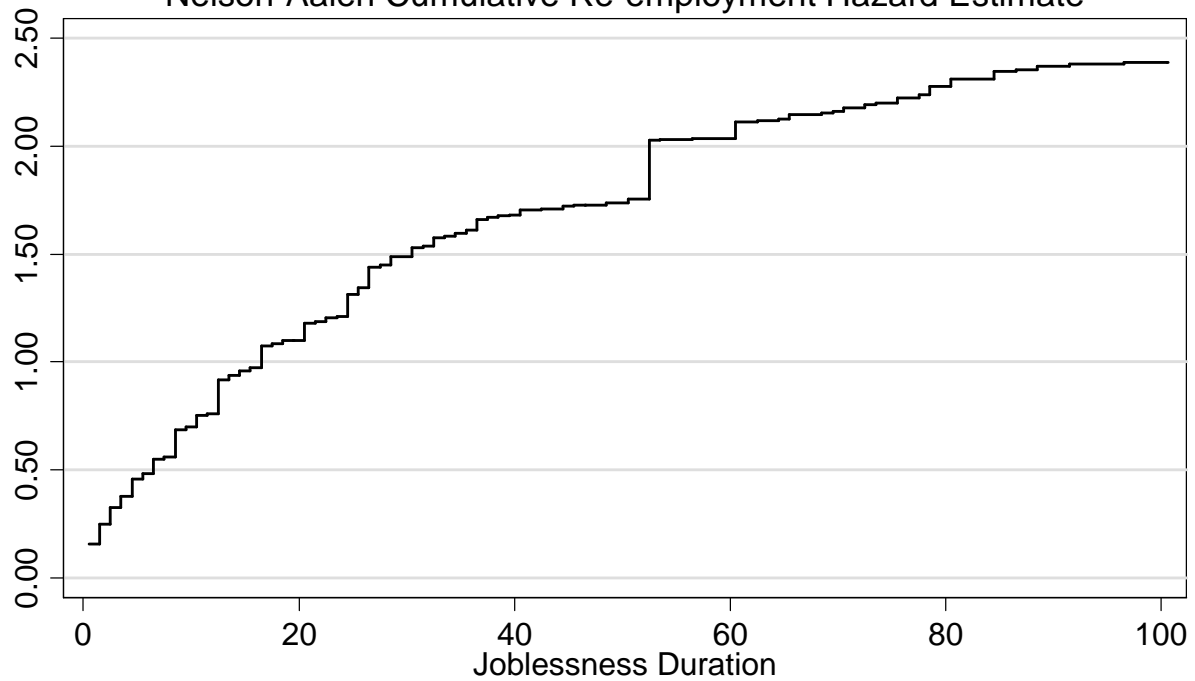
Since  $\hat{S}(0) = 1$  we then have

$$\hat{S}(t) = \prod_{s \leq t} \left\{ 1 - \left( \frac{dN^*(s)}{R(s)} \right) \right\} = \prod_{t_i \leq t} \left\{ 1 - \left( \frac{1}{R(t_i)} \right) \right\}. \quad (13)$$

Before turning to models with covariates, we present an example using joblessness spell data from the February 1996 Current Population Survey's Displaced Worker Supplement (CPS-DWS). In the CPS-DWS workers who have been displaced from a job in the previous three years are asked how many weeks it took before they were re-employed. Joblessness duration data are right censored if the spell was ongoing at the time of the survey. For convenience we also censor all spells after 100 weeks. The Nelson-Aalen estimate of the integrated or cumulative hazard function is presented in Figure 3. Since the integrated hazard is discontinuous it is not possible to directly estimate the baseline hazard. However, applying kernel smoothing techniques an estimate can be derived. This is presented in Figure 4.

Figure 3

Nelson-Aalen Cumulative Re-employment Hazard Estimate



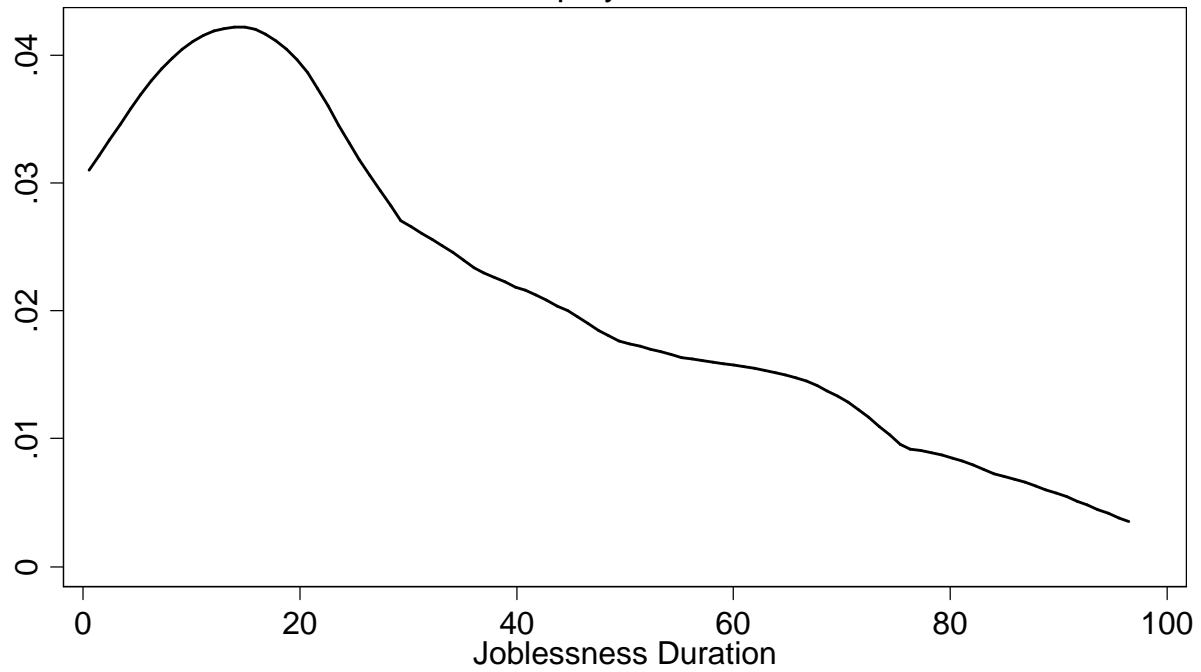
Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.



Figure 4

Smoothed Re-employment Hazard Estimate



Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

#### **4. Parametric Methods for Continuous-time Data with Covariates**

##### **i) Time-Constant Covariates**

Usually in economics we are interested in hazard functions conditional on a set of covariates or regressors. When these do not change over time, then we simply define the hazard conditional on the covariates. Thus the conditional hazard is

$$\lambda(t; \mathbf{x}) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, \mathbf{x})}{\Delta t} \quad (14)$$

where  $\mathbf{x}$  is a vector of explanatory variables. All of the formulas introduced in section 2 above continue to hold provided the c.d.f. and density are defined conditional on  $\mathbf{x}$ . Often we are interested in the partial effects of the  $\mathbf{x}_j$  on  $\lambda(t; \mathbf{x})$ , which are defined as the partial derivatives for continuous  $\mathbf{x}_j$  and differences for discrete  $\mathbf{x}_j$ . While the durations defined by (14) refer to some “internal” time until the occurrence of an event, the impact of calendar time can be modeled by incorporating suitable covariates into  $\mathbf{x}$ .

An especially important class of models with time-constant regressors consists of the proportional hazards model. A proportional hazards model can be written as  $\lambda(t; \mathbf{x}) = \rho(\mathbf{x})\lambda_0(t)$  where  $\rho(\mathbf{x}) > 0$  is a non-negative function of  $\mathbf{x}$  and  $\lambda_0(t) > 0$  is called the baseline hazard. The baseline hazard is common to all individuals in the population; individual hazards differ proportionately based on a function  $\rho(\mathbf{x})$  of observed covariates. Typically,  $\rho(\mathbf{x})$  is

parameterized as  $\rho(x) = \exp(\mathbf{x}'\boldsymbol{\beta})$

where  $\boldsymbol{\beta}$  is a vector of parameters.

## **ii) Time-Varying Covariates**

Studying hazard functions is more complicated when we wish to model the effects of time-varying covariates on the hazard function. For one thing it makes no sense to specify the distribution of the duration  $T$  conditional on the covariates at only one time period. Nevertheless, we can still define the appropriate conditional probabilities that lead to a conditional hazard function.

Let  $\mathbf{x}(t)$  denote the vector of regressors at time  $t$ . For  $t \geq 0$ , let  $\mathbf{X}(t)$ ,  $t \geq 0$  denote the covariate path through time  $t$ :  $\mathbf{X}(t) = \{\mathbf{x}(s): 0 \leq s \leq t\}$ . We define the conditional hazard function at time  $t$  by

$$\lambda(t; \mathbf{X}(t)) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t, \mathbf{X}(t + \Delta t))}{\Delta t}. \quad (15)$$

The proportional hazard form is commonly used when covariates are time-varying:

$$\lambda(t; \mathbf{X}(t)) = \rho(\mathbf{x}(t))\lambda_0(t). \text{ Usually } \rho(t) = \exp[\mathbf{x}(t)'\boldsymbol{\beta}].$$

Below, we shall focus on techniques primarily for flow sampling. With flow sampling, the sample consists of individuals who enter the state at some point during the interval  $[0, t^0]$  and we record the length of time each individual is in the initial state. Flow data are usually subject to

**right censoring.** That is, after a certain amount of time ( $t^0$ ), we stop following the individuals in the sample, which we must do in order to analyze the data. For individuals who have completed their spells in their initial state we observe the exact duration. But for those still in the initial state, we only know that the duration lasted at least as long as the tracking period.

### **iii) Maximum Likelihood Estimation with Censored Data**

For a random draw  $i$  from the population, let  $e_i \in [0, t^0]$  denote the time at which individuals  $i$  enters the initial state (the starting time), let  $t_i^*$  denote the length of time in the initial state (the duration), and let  $\mathbf{x}_i$  denote the vector of observed covariates. We assume that  $t_i^*$  has a continuous conditional density  $f(t|\mathbf{x}_i; \boldsymbol{\theta})$ ,  $t \geq 0$ , where  $\boldsymbol{\theta}$  is the vector of unknown parameters.

Without right censoring we would observe  $(e_i, \mathbf{x}_i)$  and estimation would proceed by conditional maximum likelihood estimation. To account for right censoring we assume that the observed duration  $t_i$  is obtained as  $t_i = \min(t_i^*, c_i)$  where  $c_i$  is the censoring time for individual  $i$ . In some cases,  $c_i$  is constant across  $i$ . For example if you were to track all individuals whose duration starts at the same calendar time and track them for up to 2 years then the common censoring time would be 104 weeks. We assume that, conditional on the covariates, the true duration distribution is independent of the starting point  $e_i$  and the censoring time  $c_i$ .

$$H(t_i^* | \mathbf{x}_i, e_i, c_i) = H(t_i^* | \mathbf{x}_i) \quad (16)$$

where  $H(\cdot)$  denotes the conditional distribution. Under assumption (16), the distribution of  $t_i^*$  given  $(\mathbf{x}_i, e_i, c_i)$  does not depend on  $(e_i, c_i)$ . Therefore, if the duration is not censored, the density

of  $t_i$  given  $(\mathbf{x}_i, e_i, c_i)$  is simply  $f(t|\mathbf{x}_i; \theta)$ . The probability that  $t_i$  is censored is

$$P(t_i^* \geq c_i | \mathbf{x}_i) = 1 - F(c_i | \mathbf{x}_i; \theta)$$

Let  $d_i$  be a complete spell indicator ( $d_i = 1$  if uncensored,  $d_i = 0$  if censored), the conditional likelihood for observation  $i$  can be written as

$$f(t_i | \mathbf{x}_i; \theta)^{d_i} [1 - F(t_i | \mathbf{x}_i; \theta)]^{(1-d_i)}$$

For a random sample of size  $N$  the maximum likelihood estimator of  $\theta$  is obtained by maximizing

$$\sum_{i=1}^N \{d_i \log[f(t_i | \mathbf{x}_i; \theta)] + (1 - d_i) \log[1 - F(t_i | \mathbf{x}_i; \theta)]\}$$

For example, the Weibull distribution with covariates has conditional density

$$f(t_i | \mathbf{x}_i; \theta) = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \alpha t^{\alpha-1} \exp[-\exp(\mathbf{x}_i' \boldsymbol{\beta}) t^\alpha]$$

where  $\mathbf{x}_i$  contains unity as its first element for all  $i$ .

#### **iv) Unobserved Heterogeneity**

One way to obtain more general duration models is to introduce unobserved heterogeneity into fairly simple duration models. In addition, we sometimes want to test for duration dependence conditional on observed covariates and unobserved heterogeneity.

The key assumptions used to incorporate unobserved heterogeneity are that:

- 1) Unobserved heterogeneity is independent of the observed covariates.
- 2) Unobserved heterogeneity distribution is known up to a finite number of parameters
- 3) Unobserved heterogeneity enters the hazard function in a multiplicative fashion.

Before moving to a more general framework we will consider the model by Lancaster (1979).

For a random draw  $i$  from the population it is assumed that the hazard function has the Weibull form conditional on the observed covariates  $\mathbf{x}_i$  and unobserved heterogeneity  $v_i$ :

$$\lambda(t; \mathbf{x}_i, v_i) = v_i \exp(\mathbf{x}_i' \boldsymbol{\beta}) \alpha t^{\alpha-1} \quad (17)$$

where  $x_{i1} \geq 1$  and  $v_i > 0$ . To identify the parameters  $\alpha$  and  $\beta$  we need to normalize the distribution of  $v_i$  so that  $E(v_i) = 1$ . In Lancaster (1979), it was assumed that the distribution of  $v_i = \text{Gamma}(\gamma, \gamma)$  so that  $E(v_i) = 1$  and  $\text{Var}(v_i) = 1/\gamma$ .

In the general case where the c.d.f. of given  $(\mathbf{x}_i, v_i)$  is  $F(t | \mathbf{x}_i, v_i, \gamma)$  we obtain the distribution of  $t_i^*$  given  $\mathbf{x}_i$  by integrating out the unobserved effect. Because  $v_i$  and  $\mathbf{x}_i$  are independent, the c.d.f. of  $t_i^*$  given  $\mathbf{x}_i$  is

$$G(t | \mathbf{x}_i; \boldsymbol{\theta}, \boldsymbol{\rho}) = \int_0^\infty F(t | \mathbf{x}_i, v_i; \boldsymbol{\theta}) k(v; \boldsymbol{\rho}) dv$$

where it is assumed that the density of  $v_i$ ,  $k(\cdot)$  is assumed to be continuous and depend on the unknown vector of parameters  $\theta$ .

With censoring and flow data we should assume

$$H(t_i^* | \mathbf{x}_i, v_i, e_i, c_i) = H(t_i^* | \mathbf{x}_i, v_i) \text{ and } K(v_i | \mathbf{x}_i, e_i, c_i) = K(v_i)$$

Suppose that the unobserved heterogeneity distribution has a gamma distribution and

$$\lambda(t; \mathbf{x}_i, v_i) = v_i \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta})$$

Then,

$$F(t | \mathbf{x}_i, v_i) = 1 - \exp \left[ -v_i \exp(\mathbf{x}_i' \boldsymbol{\beta}) \int_0^t \lambda_0(s) ds \right] = 1 - \exp \left[ -v_i \exp(\mathbf{x}_i' \boldsymbol{\beta}) \Lambda(t) \right]$$

$$\text{where } \Lambda(t) \equiv \int_0^t \lambda_0(s) ds.$$

Now the density of  $v_i$  is

$$k(v) = \delta^\delta v^{\delta-1} \exp(-\delta v) / \Gamma(\delta)$$

where  $\text{Var}(v_i) = 1/\delta$  and  $\Gamma(\cdot)$  is the Gamma Function. Thus,

$$G(t | \mathbf{x}_i; \theta, \rho) = 1 - \int_0^\infty \exp(-v_i \exp(\mathbf{x}_i' \boldsymbol{\beta}) \Lambda(t)) \delta^\delta v^{\delta-1} \exp(-\delta v) / \Gamma(\delta) dv$$

$$= 1 - \left( 1 + \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}) \Lambda(t)}{\delta} \right)^{-\delta}.$$

Why would we introduce heterogeneity when the heterogeneity is assumed to be independent of the observed covariates? In many instances in economics, such as job search theory, we are interested in testing for duration dependence conditional on the observed and unobserved heterogeneity, where the unobserved heterogeneity enters the hazard multiplicatively. As shown by Lancaster (1979), ignoring multiplicative heterogeneity in the Weibull model results in asymptotically underestimating  $\rho$ . Therefore, we could very well conclude that there is negative duration dependence conditional on  $\mathbf{x}$ , whereas there is no duration dependence conditional on  $\mathbf{x}$  and  $v$ .

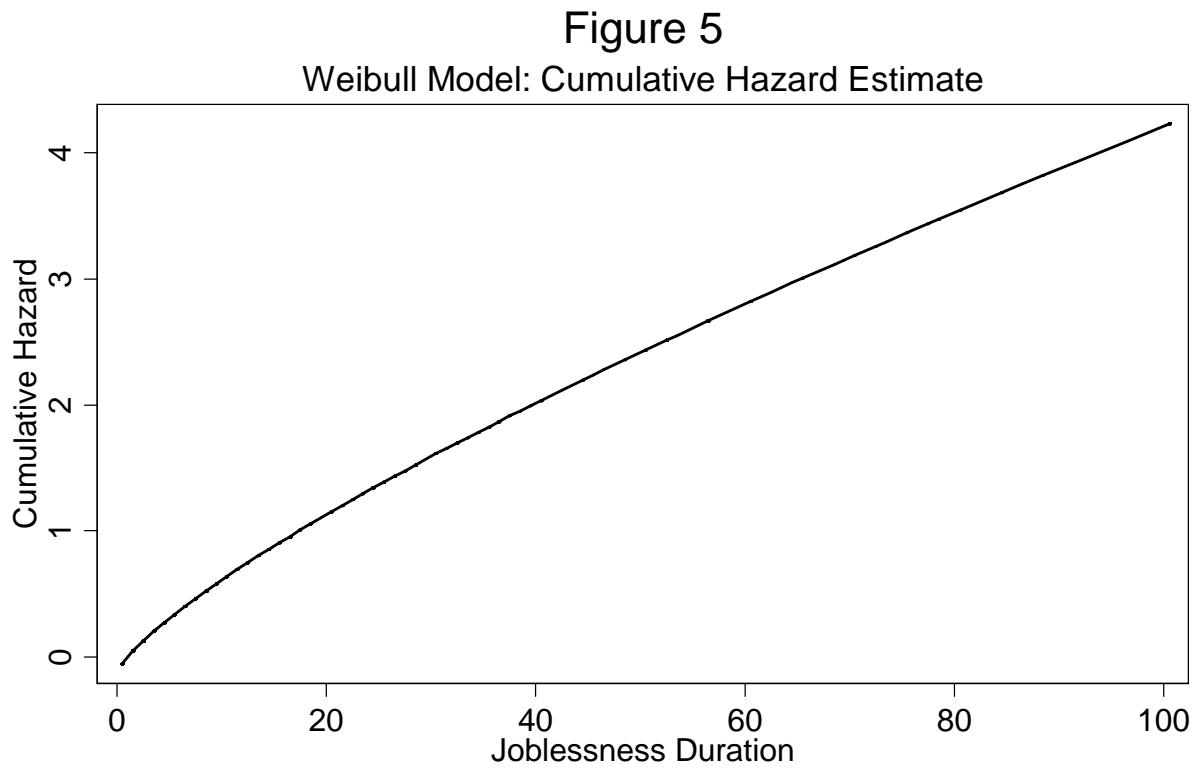
Returning to our example using the CPS-DWS we estimate both Weibull and Weibull-Gamma models controlling for a number of covariates including the weekly benefit amount an individual is qualified to receive (WBA) and an indicator for UI receipt (UI) and the interaction of the two (UI\*WBA).<sup>1</sup> Figures 5 and 6 display the estimated cumulative hazard and survivor function, respectively, when the covariates are set to their sample means. To investigate the impact of UI receipt, Figure 7 portrays the difference in the estimated survivor function for a UI recipient and UI non-recipient who both qualify for weekly benefits of \$200 per week, and whose remaining covariates are fixed at their sample means. As can be seen from the figure, the survivor function

---

<sup>1</sup> Controls for gender, race marital status, age, education, immigrant status, region of country, industry of lost job, tenure in lost job, bluecollar-whitecollar status, reason for displacement, weekly wage in lost job and year of displacement were also included. The sample includes only those who are imputed to be eligible for UI benefits.



of the non-recipient decreases much more rapidly indicating that they find jobs more quickly than UI recipients. Figures 8 through 10 present graphs for the Weibull-Gamma model.

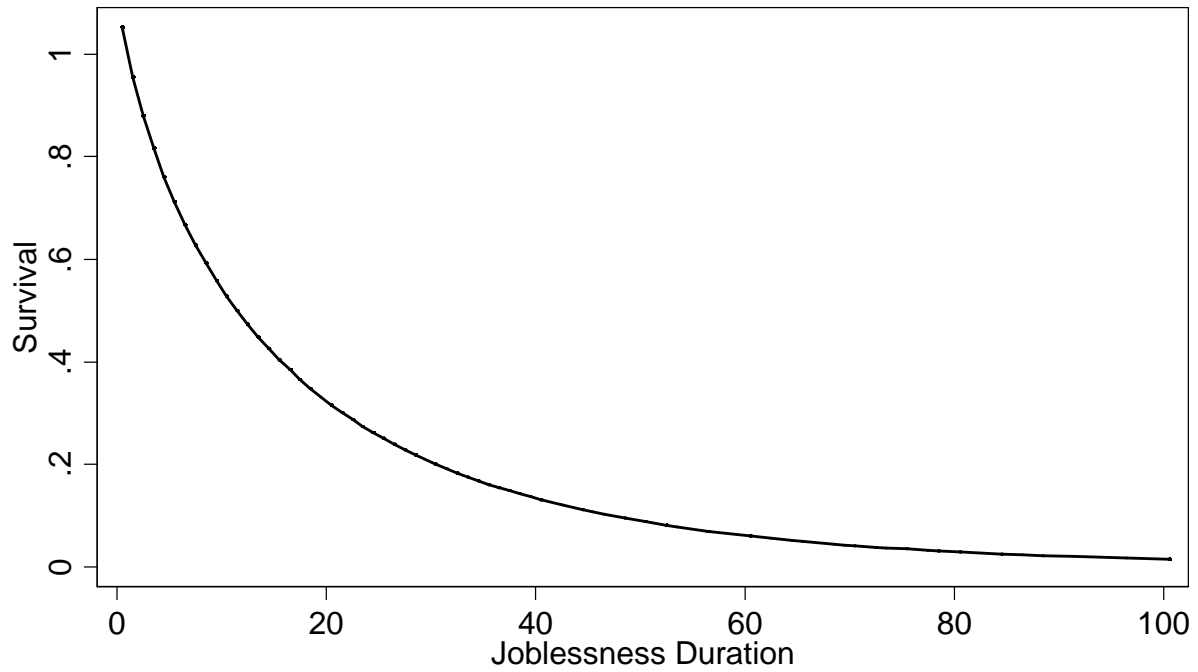


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 6

Weibull Model: Survivor Function Estimate

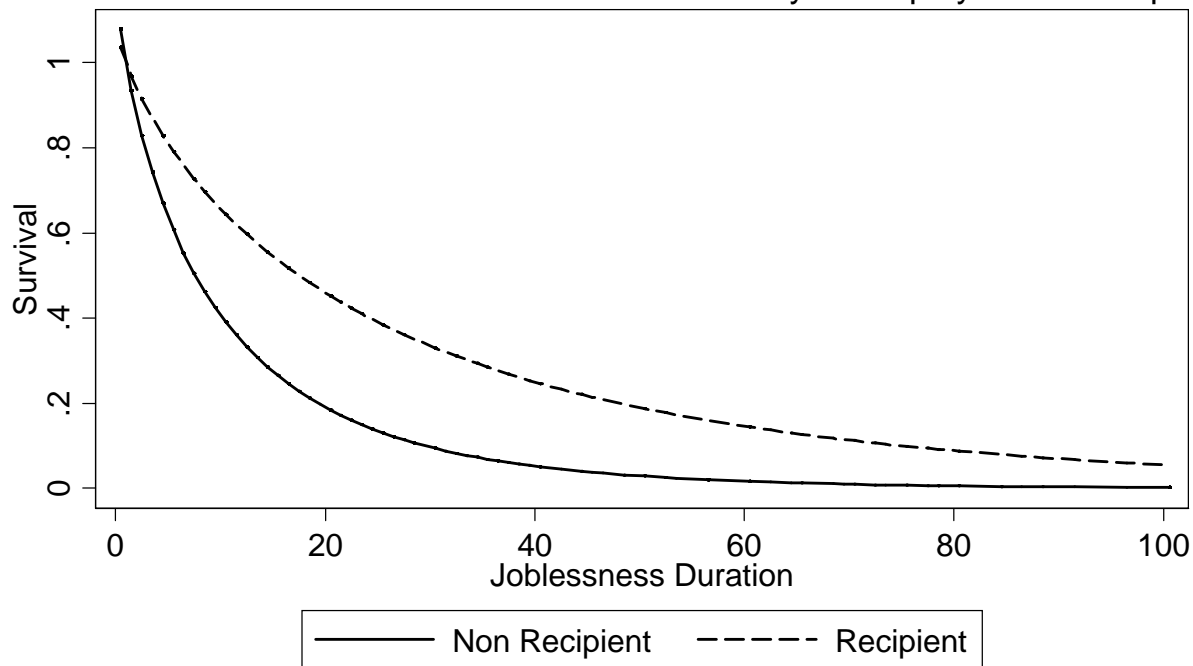


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 7

Weibull Model: Survivor Function Estimate by Unemployment Receipt

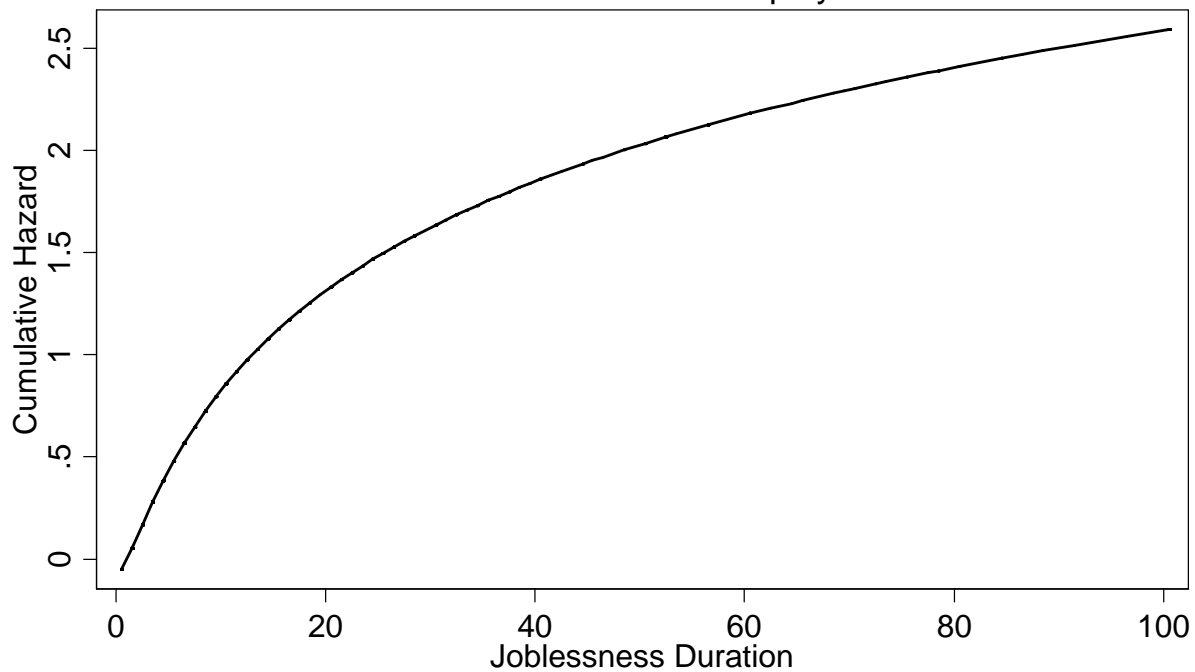


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 8

Weibull-Gamma Model: Cumulative Re-employment Hazard Estimate

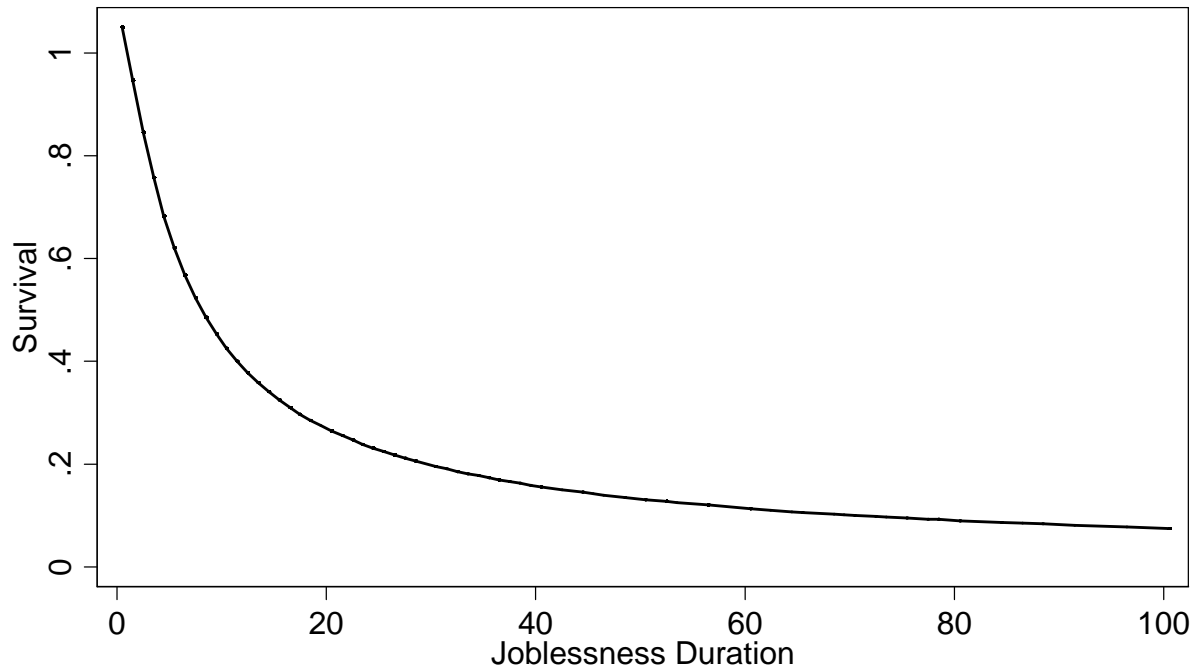


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 9

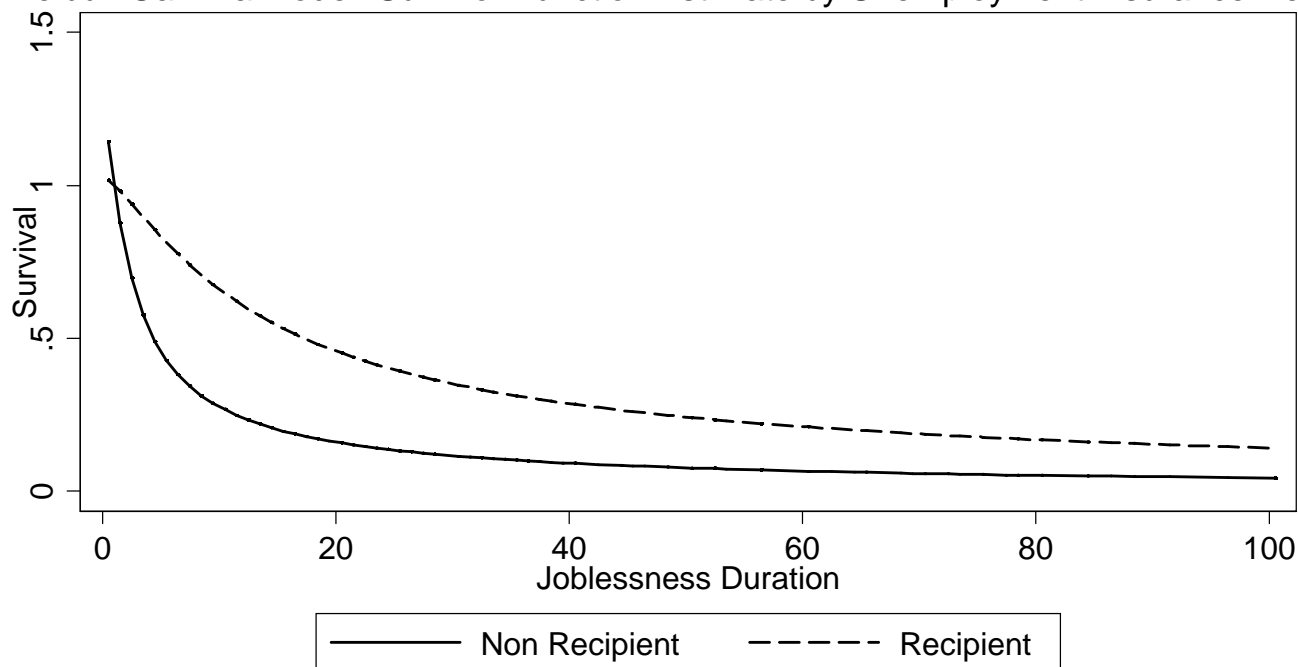
Weibull-Gamma Model: Survivor Function Estimate



Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

**Figure 10**  
Weibull-Gamma Model: Survivor Function Estimate by Unemployment Insurance Rece



Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp., College Station, TX.

## **5. The Cox Regression Model**

### **i) Data with no ties:**

The models presented above impose parametric assumptions on the baseline hazard function. In many instances economic theory provides little help in identifying a particular parametric class. However, if the true baseline hazard function does not belong to the assumed parametric class of functions, estimates will generally be biased. Cox (1972, 1975) developed a technique for obtaining estimates of the  $\beta$  without imposing any parametric form on the baseline hazard. This technique is referred to as Cox regression. The model was developed with

continuous time duration data in which the probability of two durations ending at the same time equals zero. In most applications however, duration data is grouped to some extent and the model has been modified to accommodate “ties” in the data. First, however, we will look at the case of “no ties.”

The Cox regression model assumes that the conditional hazard function follows a proportional hazards model:

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta})$$

The benefit of the Cox regression method is that it makes no assumptions about the form of the baseline hazard function  $\lambda_0(t)$ . In fact the estimation method “partials out” the baseline hazard so that it doesn’t appear in the maximand; the only parameters that appear are the regression coefficients. Thus, Cox regression is a semi-parametric estimation method.

Cox regression estimation relies on forming a risk pool or risk set at each failure time in the data. The risk set at failure time  $t_j$  includes all individuals,  $i$ , with  $c_i$  greater than or equal to  $t_j$ . Thus at  $t_j$  an individual is in the risk set if the event has not occurred before that time or they have not been censored. The partial likelihood function is then constructed by considering the conditional probabilities of failure at each failure time.

For example suppose that there are 5 observations in the data such that

Obs.	$t_i$	$d_i$	$x_i$
1.	3	1	2
2.	5	0	2
3.	6	1	1
4.	9	1	0
5.	11	1	1

Let  $R(t_j)$  be the risk set at the (ordered) failure time  $t_j$ . Thus,  $R(3) = \{1,2,3,4,5\}$ ,  $R(6) = \{3,4,5\}$ ,  $R(9) = \{4,5\}$  and  $R(11)=\{5\}$ . At each event time  $t_j$ , we consider the conditional probability that the event occurs for the particular observation among those observations remaining in the risk set “just before”  $t_j$ , conditional on one event occurring at  $t_j$ . For any observation,  $i$ , in the risk set at time  $t_j$  the probability of failure at  $t_j$  approximately equals  $\varphi(t_j; \mathbf{x}_i)$ . Thus, if observation  $j$  fails at time  $t_j$ , then the conditional probability of observing  $j$  equals

$$\frac{\exp(\mathbf{x}'_j \boldsymbol{\beta}_j)}{\sum_{i \in R(t_j)} \exp(\mathbf{x}'_i \boldsymbol{\beta}_i)} \text{ or simply } \frac{\exp(\mathbf{x}'_j \boldsymbol{\beta}_j)}{\sum_{i \in R_j} \exp(\mathbf{x}'_i \boldsymbol{\beta}_i)} \text{ where } R_j \equiv R(t_j).$$

An alternative way to think about the construction of the partial likelihood is to think of drawing balls from an urn at each failure time. A ball is included in the urn for individual  $i$  at time  $t_j$  only if individual  $i$  has not been censored or has not failed before time  $t_j$ . Instead of equal probabilities the relative probability of drawing the ball associated with individual  $i$  equals

$$\exp(\mathbf{x}'_i \boldsymbol{\beta}) . \text{ Thus probability of drawing individual } j \text{ equals } \frac{\exp(\mathbf{x}'_j \boldsymbol{\beta}_j)}{\sum_{i \in R(t_j)} \exp(\mathbf{x}'_i \boldsymbol{\beta}_i)} .$$

The partial likelihood is formed by the product of these conditional probabilities over all failure times:

$$PL(\boldsymbol{\beta}) = \prod_{j=1}^K \left( \frac{\exp(\mathbf{x}'_j \boldsymbol{\beta})}{\sum_{i \in R_j} \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right)$$

or

$$\log PL(\boldsymbol{\beta}) = \sum_{j=1}^K \mathbf{x}'_j \boldsymbol{\beta} - \log \left( \sum_{i \in R_j} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right) \quad (18)$$



Estimates are obtained by maximizing (18) with respect to  $\boldsymbol{\beta}$ . Let  $\hat{\boldsymbol{\beta}}$  denote the value of  $\boldsymbol{\beta}$  that maximizes (18). Then, the first order conditions for a maximum state that the vector  $\hat{\boldsymbol{\beta}}$  must satisfy:

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \sum_{j=1}^K \left( \mathbf{x}_j - \frac{\sum_{i \in R_j} \mathbf{x}_i \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})}{\sum_{i \in R_j} \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})} \right) = \mathbf{0}$$

The vector function  $\mathbf{s}$  is usually referred to as the score function. From our discussion of counting processes it is clear that

$$M_i(t) = \int_0^t Y(s)Z(s)dN_i(s) - \int_0^t Y(s)Z(s)\lambda_0(s)\exp(\mathbf{x}_i' \boldsymbol{\beta})ds$$

is a martingale. Defining the predictable function  $H_j(t, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  as

$$H_j(t, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \mathbf{x}_j - \frac{\sum_{i \in R_j} \mathbf{x}_i \exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{i \in R_j} \exp(\mathbf{x}_i' \boldsymbol{\beta})}$$

we have

$$\begin{aligned}
s_j(\boldsymbol{\beta}) &= \sum_{j=1}^n \int_0^\infty H_j(s, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) dN_j(s) \\
&= \sum_{j=1}^n \int_0^\infty H_j(s, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) \left\{ dM_j(s) - Y_j(s) Z_j(s) \lambda_0(s) \exp(\mathbf{x}_j' \boldsymbol{\beta}) \right\} \\
&= \sum_{j=1}^n \int_0^\infty H_j(s, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) dM_j(s) - \\
&\quad \sum_{j=1}^n \int_0^\infty H_j(t, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) \lambda_0(t) \exp(\mathbf{x}_j' \boldsymbol{\beta}) \\
&= \sum_{j=1}^n \int_0^\infty H_j(s, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) dM_j(s)
\end{aligned} \tag{19}$$

since

$$\begin{aligned}
&\sum_{j=1}^n \int_0^\infty H_j(s, \boldsymbol{\beta}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) Y_j(s) Z_j(s) \lambda_0(s) \exp(\mathbf{x}_j' \boldsymbol{\beta}) ds \\
&= \sum_{j=1}^n \int_0^\infty \left( \mathbf{x}_j - \frac{\sum_{i \in R_j} \mathbf{x}_i \exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{i \in R_j} \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right) Y_j(s) Z_j(s) \lambda_0(s) \exp(\mathbf{x}_j' \boldsymbol{\beta}) ds \\
&= \int_0^\infty \left( \sum_{j=1}^n Y_j(s) Z_j(s) \mathbf{x}_j \exp(\mathbf{x}_j' \boldsymbol{\beta}) - \frac{\exp(\mathbf{x}_j' \boldsymbol{\beta}) \sum_{i \in R_j} \mathbf{x}_i \exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{i \in R_j} \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right) \lambda_0(s) ds \\
&= \int_0^\infty \left( \sum_{i \in R_j} \mathbf{x}_j \exp(\mathbf{x}_j' \boldsymbol{\beta}) - \sum_{i \in R_j} \frac{\exp(\mathbf{x}_j' \boldsymbol{\beta}) \sum_{i \in R_j} \mathbf{x}_i \exp(\mathbf{x}_i' \boldsymbol{\beta})}{\sum_{i \in R_j} \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right) \lambda_0(s) ds = 0.
\end{aligned}$$

Using (19) and appealing to results from laws of large numbers and Martingale Central

Limit Theory, it can be shown that the estimates are both consistent and  $\sqrt{n}$  - asymptotically normal with the variance-covariance matrix equal to

$$\left[ -E \left( -\frac{\partial^2 \log(PL(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) \right]^{-1}$$

which can be consistently estimated by

$$\left[ \sum_{j=1}^k -\frac{\partial^2 \log(PL_j(\hat{\boldsymbol{\beta}}))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right]^{-1} \text{ or } \left[ \sum_{j=1}^k \frac{\partial \log(PL_j(\hat{\boldsymbol{\beta}}))}{\partial \boldsymbol{\beta}} \frac{\partial \log(PL_j(\hat{\boldsymbol{\beta}}))}{\partial \boldsymbol{\beta}'} \right]^{-1}$$

where the vector  $\hat{\boldsymbol{\beta}}$  denotes the vector of Cox regression estimates and

$$PL_j(\hat{\boldsymbol{\beta}}) = \frac{\exp(\mathbf{x}'_j \hat{\boldsymbol{\beta}})}{\sum_{i \in R_j} \exp(\mathbf{x}'_i \hat{\boldsymbol{\beta}})}.$$

## ii) Data with Ties

With ties in the data the exact partial likelihood becomes more complex although there are some approximation methods that reduce the complexity and perform well as long as the number of ties are “small”. Suppose at each event time  $j$ ,  $d_j$  events occur and let  $\mathbf{D}_j$  be the set of individuals for which the event occurs at time  $j$ . Returning to the urn analogy then instead of drawing one ball from the urn at time  $j$ , we draw  $d_j$  balls without replacement. For a particular sequence  $s = \{j(1), \dots, j(d_j)\}$  of draws the probability of observing that sequence equals

$$\prod_{q=1}^{d_j} \frac{\exp(\mathbf{x}'_{j(q)} \boldsymbol{\beta}_{j(q)})}{\sum_{i \in R(t_j) - \{j(1), \dots, j(q-1)\}} \exp(\mathbf{x}'_i \boldsymbol{\beta}_i)} . \quad (20)$$

Thus, the probability equals

$$\frac{1}{d_j!} \sum_{s \in P(\mathbf{D}_j)} \prod_{q=1}^{d_j} \frac{\exp(\mathbf{x}'_{j(q)} \boldsymbol{\beta}_{j(q)})}{\sum_{i \in R(t_j) - \{j(1), \dots, j(q-1)\}} \exp(\mathbf{x}'_i \boldsymbol{\beta}_i)}$$

where  $P(\mathbf{D}_j)$  represents the set of permutations of the indices of the  $\mathbf{D}_j$  individuals who fail at time  $t_j$ . Since the construction of the exact partial likelihood with ties can be quite complex, several approximations have been suggested. Perhaps the most well know is that by Breslow(1974) who essentially substitutes sampling with replacement for sampling without replacement in the urn analogy.

Define

$$\mathbf{s}_j = \sum_{k \in \mathcal{D}_k} \mathbf{x}_k .$$

Then the Breslow approximation to the partial likelihood equals:

$$PL_B(\boldsymbol{\beta}) = \prod_{j=1}^K \left( \frac{\exp(\mathbf{s}'_j \boldsymbol{\beta})}{\left[ \sum_{i \in R_j} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right]^{d_j}} \right) .$$

An alternative approximation by Efron (1977) adjusts the denominator of the Breslow approximation to the partial likelihood by subtracting a term for the number of balls drawn from the urn. But instead of deducting the probability weights for the actual balls drawn and then averaging over all permutations Efron (1977) deducts the average probability weight where the

average is over all  $d_j$  in  $\mathbf{D}_j$ . The Efron (1977) approximation to the partial likelihood equals:

$$PL_E(\boldsymbol{\beta}) = \prod_{j=1}^K \left( \frac{\exp(\mathbf{s}'_j \boldsymbol{\beta})}{\prod_{k=1}^{d_j} \left[ \sum_{i \in R_j} \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \frac{k-1}{d_j} \sum_{i \in \mathcal{D}_j} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right]} \right).$$

Since the Cox regression partial likelihood eliminates the baseline hazard function it does not produce estimates of the baseline hazard function. The estimates of  $\boldsymbol{\beta}$ , however, can be used to estimate the cumulative baseline hazard,  $\Lambda_0(t)$ , using an estimator that weights the data using  $\exp(\mathbf{x}'_i \hat{\boldsymbol{\beta}})$ :

$$\hat{\Lambda}_0(t) = \int_0^t \frac{J(s)}{\sum_{i=1}^n Y_i(s) Z_i(s) \exp(\mathbf{x}'_i \hat{\boldsymbol{\beta}})} dN^*(s).$$

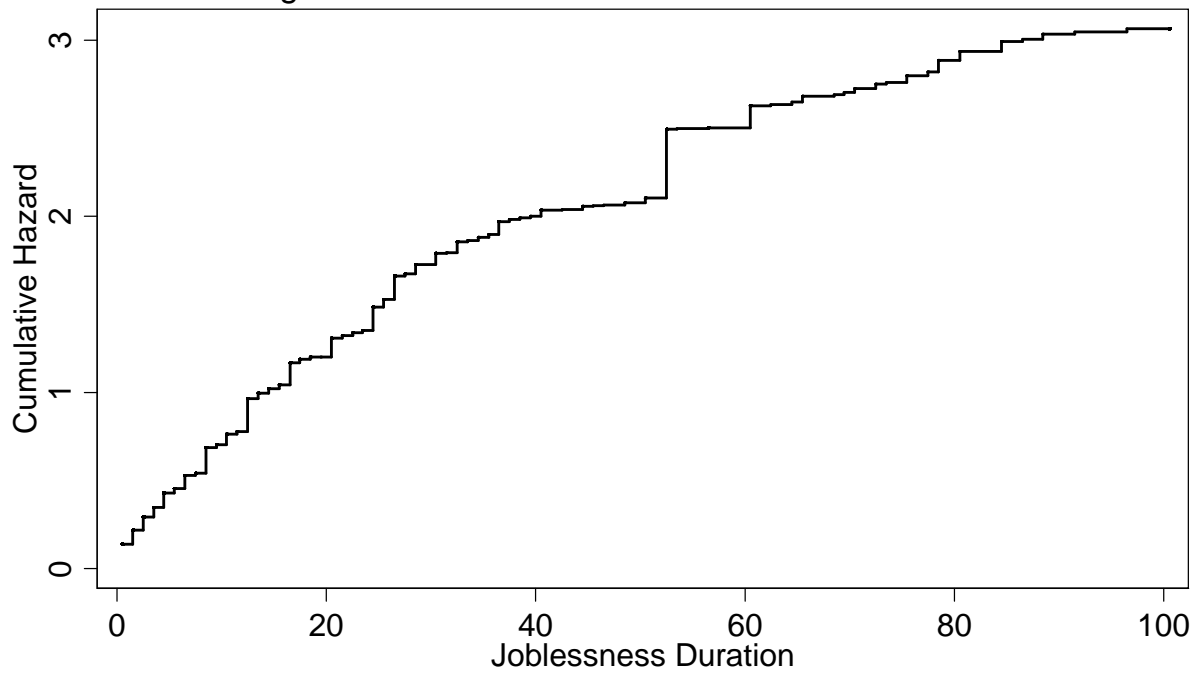
An estimate of the baseline survivor function,  $S_0(t)$ , then equals:

$$\hat{S}_0(t) = \exp(-\hat{\Lambda}_0(t)).$$

Returning to our joblessness example presented above, Figures 11 and 12 present estimates of the cumulative hazard and survivor functions when covariates are fixed at the sample mean while Figure 13 presents estimates of the survivor function for a UI recipient and non-recipient who qualify for \$200 per week in benefits and whose other covariates are fixed at the sample mean.

Figure 11

Cox Regression Model: Cumulative Hazard Function Estimate

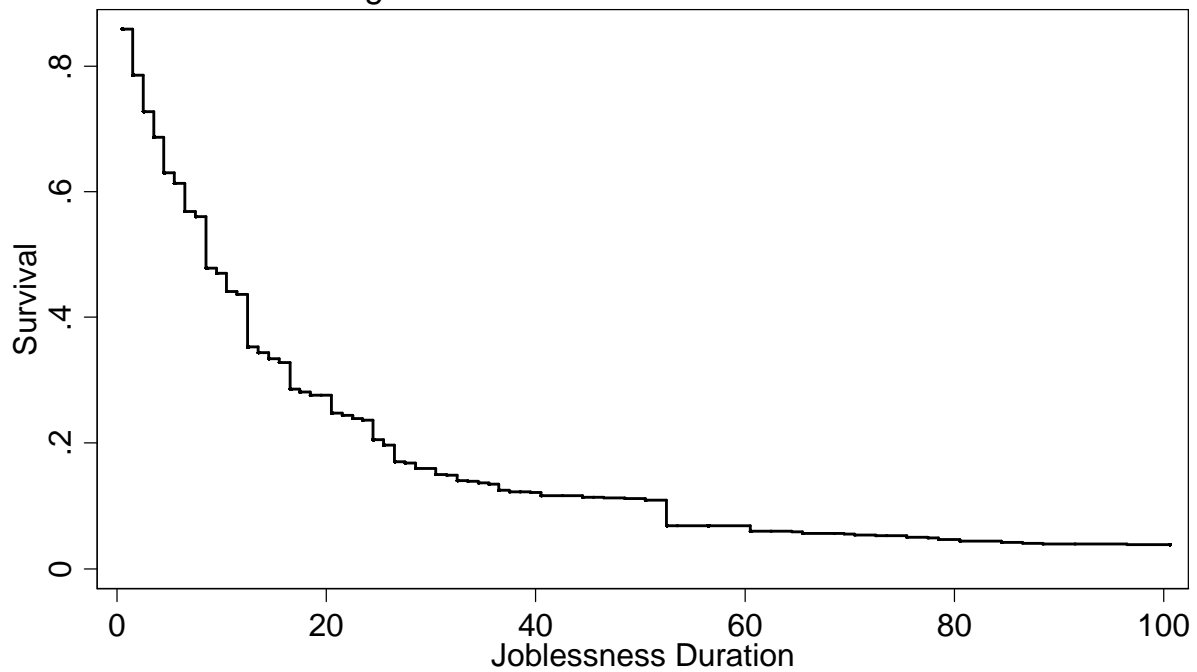


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 12

Cox Regression Model: Survivor Function Estimate

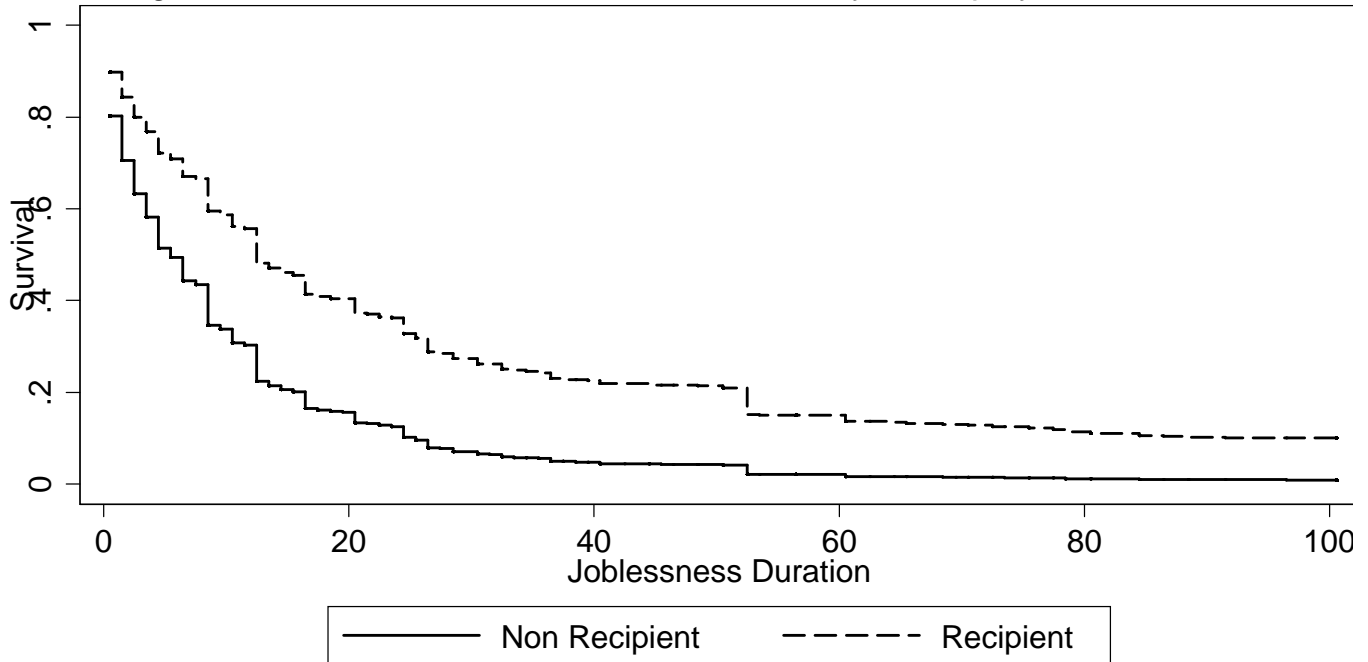


Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

Figure 13

Cox Regression Model: Survivor Function Estimate by Unemployment Insurance Recipient



Source: February 1996 Current Population Survey

Graph produced by Stata 8.2  
Stata Corp. College Station, TX.

### iii) Stratified Cox Regression

In some circumstances the proportional hazards assumption may be inappropriate. If the suspect variable is a categorical variable, then one can relax the proportional hazards assumption for that variable by estimating a stratified model. Suppose the variable  $w$  has  $H$  categories and that, a priori, it you suspect that the hazards are non-proportional across the  $H$  categories. Then the stratified Cox regression involves maximizing the partial likelihood function:



$$SPL(\beta) = \prod_{h=1}^H \prod_{j^h=1}^{K^h} \left( \frac{\exp(\mathbf{x}'_{j^h} \beta)}{\sum_{i \in R_{j^h}} \exp(\mathbf{x}'_i \beta)} \right)$$

where  $\mathbf{x}$  includes all other variables except  $w$ .

## **6. Discrete-time Duration Data**

### **i) Time-Constant Covariates**

Most duration data available in economics is grouped. That is durations are only known to fall into a certain time intervals, such as weeks, month, or even years. For example, unemployment duration data are typically grouped into weeks.

One econometric approach that is taken when analyzing grouped duration data summarizes the information on staying in the initial state or exiting that state in each time interval using a sequence of binary responses. Essentially, we have a panel data set where the duration of an individual determines a vector of binary responses. These in conjunction with the covariates can be thought of as creating an unbalanced panel where the number of observations per individual equals  $K_i = \min(D_i, C_i)$  where  $D_i$  equals the number of periods until the event occurs and  $C_i$  equals the number of periods until the observation is (right) censored. In addition to allowing us to treat grouped durations, the panel data approach has at least two advantages. First, in a proportional hazard specification it leads to simple methods for estimating flexible hazard functions. Second, because of the sequential nature of the data, time varying covariates are easily introduced.

Through out this discussion we will assume flow sampling. We divide the time line into

$M + 1$  intervals,  $[0, k_1)$ ,  $[k_1, k_2)$ ,  $?$ ,  $[k_{M-1}, k_M)$ ,  $[k_M, ?)$ , where  $k_m$  are known constants. For example, we might have  $k_1 = 1$ ,  $k_2 = 2$ ,  $k_3 = 3$ , and so on, but unequally spaced intervals are also feasible. The last interval is chosen so that any duration falling into it is censored at  $k_M$ : no observed durations are greater than  $k_M$ .

For a random draw from the population, Let  $y_m$  represent a binary indicator equal one if the event occurs in the  $m^{\text{th}}$  interval and zero otherwise. For each person  $i$ , we observe  $(y_{i1}, \dots, y_{iK_i})$  which is an unbalanced panel data set of length  $K_i$ . Note that the string of binary indicators is either a sequence of all zeros or a series of zeros ending with a one where the former sequence is observed when the observation is censored and the latter sequence is observed when the series of zeros ends because an event occurred.

Let  $\delta_{ik} = I(C_i = k)$ ,  $k_i = 1, \dots, K_i$ . With time invariant covariates, each draw from the population is,  $\{(y_{i1}, \delta_{i1}), \dots, (y_{iK_i}, \delta_{iK_i}), \mathbf{x}\}$ . We assume that a parametric hazard function is specified as  $\lambda(t; \mathbf{x}, \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a vector of unknown parameters. Let  $T$  denote the time until exit from the initial state. While we do not fully observe  $T$ , either we know which interval it falls into, or we know whether it was censored in a particular interval. Thus we can compute

$$p(y_m = 0 | y_{m-1} = 0, \delta_{m-1} = 0, \mathbf{x},) , \quad p(y_m = 1 | y_{m-1} = 0, \delta_{m-1} = 0, \mathbf{x},) , \quad m=1, \dots, M.$$

To compute these probabilities in terms of  $T$ , we assume that the duration is conditionally independent of censoring:

$T$  is independent of  $C$  given  $C$ . Thus,

$$P(y_m = 1 \mid y_{m-1} = 0, \delta_{m-1} = 0, \mathbf{x}) = P(k_{m-1} \leq T < k_m \mid T \geq k_{m-1}) = 1 - \exp\left[-\int_{k_{m-1}}^{k_m} \lambda(s; \mathbf{x}, \boldsymbol{\theta}) ds\right] \\ \equiv 1 - \alpha_m(\mathbf{x}, \boldsymbol{\theta}), \quad m=1, \dots, M \text{ where } \alpha_m(\mathbf{x}, \boldsymbol{\theta}) \equiv \exp\left[-\int_{k_{m-1}}^{k_m} \lambda(s; \mathbf{x}, \boldsymbol{\theta}) ds\right].$$

Therefore,

$$P(y_m = 1 \mid y_{m-1} = 0, \delta_{m-1} = 0, \mathbf{x},) = \alpha_m(\mathbf{x}, \boldsymbol{\theta}).$$

We can use these probabilities to construct the likelihood function. If, for observation  $i$ , uncensored exit occurs in interval  $m_i$ , the likelihood function is

$$\left[ \prod_{m=1}^{k_i-1} \alpha_m(\mathbf{x}_i, \boldsymbol{\theta}) \right] [1 - \alpha_{k_i}(\mathbf{x}_i, \boldsymbol{\theta})]. \quad (21)$$

The first term represents the probability of remaining in the initial state for the first  $k_i - 1$  intervals, and the second term is the (conditional) probability that  $T$  falls into interval  $k_i$ . If the duration is censored in interval  $k_i$ , we know only that exit did not occur in the first  $k_i - 1$  intervals, and the likelihood consists of only the first term in expression (21).

If  $d_i$  is a censoring indicator equal to one if duration  $i$  is uncensored, the log likelihood for observation  $i$  can be written as

$$\log(L_i) = \sum_{m=1}^{k_i-1} \log[\alpha_m(\mathbf{x}_i, \boldsymbol{\theta})] + d_i \log[1 - \alpha_{k_i}(\mathbf{x}_i, \boldsymbol{\theta})] \quad (22)$$

Thus, for a sample of size  $N$ , the log likelihood function is

$$\log(L) = \sum_{i=1}^N \log(L_i) = \sum_{i=1}^N \sum_{m=1}^{k_i-1} \log[\alpha_m(\mathbf{x}_i, \boldsymbol{\theta})] + d_i \log[1 - \alpha_{k_i}(\mathbf{x}_i, \boldsymbol{\theta})] \quad (23)$$

To implement conditional MLE, we must specify a hazard function. One hazard function that is popular is a piecewise-constant proportional hazard: for  $m=1, \dots, M$ ,

$$\lambda(t; \mathbf{x}, \boldsymbol{\theta}) = \gamma(\mathbf{x}, \boldsymbol{\beta}) \lambda_m, \quad k_{m-1} \leq t \leq k_m.$$

With a piecewise constant hazard and  $\gamma(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta})$  for  $m=1, \dots, M$  we have

$$\alpha_m(\mathbf{x}, \boldsymbol{\theta}) \equiv \exp[-\exp(\mathbf{x}'\boldsymbol{\beta}) \lambda_m (k_m - k_{m-1})]$$

where  $k_m$  are known constants (often  $k_m = m$ ).

Alternatively one could assume an underlying continuous baseline hazard and define

$$\alpha_{0m} = \int_{k_{m-1}}^{k_m} \lambda_0(s) ds, \quad m=1, \dots, M$$

and

$$\alpha_m(\mathbf{x}, \boldsymbol{\theta}) \equiv \exp[-\exp(\mathbf{x}'\boldsymbol{\beta}) \alpha_{0m}] . \quad (24)$$

Without covariates, maximum likelihood estimation of the  $\alpha_{0m}$  in (24) leads to a well known estimator of the survivor function. We can motivate the estimator from the representation of the survivor function as a product of conditional probabilities. For  $m=1, \dots, M$ , the survivor function at time  $k_m$  can be written as

$$S(k_m) = P(T > k_m) = \prod_{r=1}^m P(T > k_r | T > k_{r-1})$$

Now, for each  $r=1, 2, \dots, M$  let  $N_r$  denote the number of people in the risk set for interval  $r$ :  $N_r$  is the number of people who have neither left the initial state nor been censored by  $k_{r-1}$ . Therefore,  $N_1$  is the number of individuals in the initial random sample;  $N_2$  is the number of people who did not exit the initial state in the first interval, less the number of individuals censored in the first interval, and so on. Let  $E_r$  be the number of people observed to leave in the  $r^{\text{th}}$  interval. A consistent estimator of

$$P(T > k_r | T > k_{r-1}) \text{ is } \frac{(N_r - E_r)}{N_r}, \quad r=1, 2, \dots, M. \quad (25)$$

It follows from equation (25) that a consistent estimator of the survivor function at time  $k_n$  is

$$\hat{S}(k_m) = \prod_{r=1}^m \left[ \frac{(N_r - E_r)}{N_r} \right], \quad m=1, 2, \dots, M.$$

This is the discrete-time Kaplan-Meier estimator of the survivor function (at points  $k_1, k_2, \dots, k_m$ ).

We can derive this Kaplan-Meier estimator by maximizing the likelihood function

$$L = \prod_{i=1}^N \left[ \prod_{m=1}^{k_i-1} \alpha_m \right] (1 - \alpha_{k_i})^{d_i}. \quad (26)$$

with respect to  $\alpha_m$ ,  $m=1, \dots, M$  where  $d_i$  is an indicator variable that equals one if the individual spell is not censored. Taking the log of (26) gives

$$\ln(L) = \sum_{i=1}^N \left\{ \sum_{m=1}^{k_i-1} \ln(\alpha_m) \right\} + d_i \ln(1 - \alpha_{k_i})$$

and then rearranging terms yields

$$\ln(L) = \sum_{m=1}^M \left\{ \sum_{i \in S(m)} \ln(1 - \alpha_m) + \sum_{i \in D(m)} \ln(\alpha_m) \right\}$$

where  $S(m)$  denotes those individuals who survivor past  $m$  and  $D(m)$  denotes the set of individuals for which the event occurs during  $m$ . This reduces to

$$\ln(L) = \sum_{m=1}^M \left\{ S^{\#}(m) \ln(\alpha_m) + D^{\#}(m) \ln(1 - \alpha_m) \right\}$$

where  $S^{\#}(m)$  ( $D^{\#}(m)$ ) denotes the number of individuals in  $S(m)$  ( $D(m)$ ). Differentiating the log likelihood and setting it to zero yields

$$S^{\#}(m) \frac{1}{\hat{\alpha}_m} = D^{\#}(m) \frac{1}{1 - \hat{\alpha}_m}$$

or

$$(S^{\#}(m) + D^{\#}(m)) \hat{\alpha}_m = S^{\#}(m) .$$

Solving for  $\hat{\alpha}_m$  then yields

$$\hat{\alpha}_m = \frac{S^{\#}(m)}{S^{\#}(m) + D^{\#}(m)} = \frac{N_m - E_m}{N_m}$$

Now

$$\hat{S}(a_m) = \prod_{r=1}^m \hat{\alpha}_r = \prod_{r=1}^m \left( \frac{N_r - E_r}{N_r} \right)$$

which was the Kaplan Meier estimator above.

## ii) Time-Varying Covariates

For the population, let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$  denote the outcomes of the covariates in each of the  $M$  time intervals and let  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ , where we assume that the covariates are constant within the interval. In general we will let  $\mathbf{X}_t = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)$ . We assume that the hazard at time  $t$  conditional on the covariates up through time  $t$  depends only on the covariates at time  $t$ . If past values of the covariates matter, they can simply be included in the covariates at time  $t$ . The conditional independence assumption on the censoring indicators is now stated as

$$D(T = k_m | T \geq k_{m-1}, \mathbf{x}_m, \delta_m) = D(T = k_m | T \geq k_{m-1}, \mathbf{x}_m), \quad m=1, 2, \dots, M$$

This assumption allows the censoring decision to depend on the covariates during the time interval (as well as past covariates, provided they are either included in  $\mathbf{x}_m$  or do not affect the distribution of  $T$  given  $\mathbf{x}_m$ ). Under this assumption, the probability of exiting (without censoring) is

$$\begin{aligned} P(y_m = 1 | y_{m-1} = 0, \mathbf{x}_m, \delta_m = 0) &= P(k_{m-1} \leq T < k_m | T \geq k_{m-1}, \mathbf{x}_m) \\ &= 1 - \exp\left[-\int_{k_{m-1}}^{k_m} \lambda(s; \mathbf{x}_m, \theta) ds\right] \equiv 1 - \alpha_m(\mathbf{x}_m, \theta). \end{aligned} \quad (27)$$

we can use equation (27) along with  $P(y_m = 0 | y_{m-1} = 0, \mathbf{x}_m, \delta_m = 0) = \alpha_m(\mathbf{x}_m, \theta)$  to construct the log likelihood for person  $i$  as in (22) and the sample log likelihood (23).

## iii) Unobserved Heterogeneity

Unobserved heterogeneity can also be added to hazard models for grouped data. For example, adding unmeasured heterogeneity to (24) and letting  $\varphi_i = \exp(v_i)$  gives

$$\alpha_m(\mathbf{x}_m, \xi_i, \boldsymbol{\theta}) \equiv \exp[-\xi_i \exp(\mathbf{x}'_m \boldsymbol{\beta}) \alpha_{om}] \quad (28)$$

Now the survivor function associated with () equals

$$S(m; \mathbf{X}_m, \xi, \boldsymbol{\theta}) = \left[ \prod_{r=1}^m \alpha_r(\mathbf{x}_r, \xi, \boldsymbol{\theta}) \right] = \left[ \prod_{r=1}^m \exp[-\xi \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or}] \right] = \exp \left[ -\xi \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or} \right]. \quad (29)$$

Let  $\varphi$  have c.d.f.  $G(\varphi)$ . Then

$$S(m; \mathbf{X}_m, \theta) = \int \exp \left[ -\xi \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or} \right] dG(\xi).$$

Now, if  $G$  is a gamma distribution with  $E(\varphi) = 1$  and  $\text{Var}(\varphi) = \varphi^2$  then (29) becomes

$$S(m; \mathbf{X}_m, \boldsymbol{\theta}) = \left( 1 + \sigma^2 \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or} \right)^{-1/\sigma^2}. \quad (30)$$

We can then use (30) to form the log likelihood function by noting that the probability of the event ending in period  $m$  equals  $S(m-1; \mathbf{X}_{m-1}, \varphi) - S(m; \mathbf{X}_m, \varphi)$ . Letting  $d_i$  equal 1 if event occurs and 0 otherwise (censored) we have the log-likelihood function



$$\ln(L) = \sum_{i=1}^N (1 - d_i) \ln[S(k_i - 1; \mathbf{X}_{k-1}, \boldsymbol{\theta})] + d_i \ln[S(k_i - 1; \mathbf{X}_{k-1}, \boldsymbol{\theta}) - S(k_i; \mathbf{X}_k, \boldsymbol{\theta})] \ln$$

where  $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \sigma^2)'$  with  $\boldsymbol{\alpha}' = (\alpha_{01}, \alpha_{02}, \dots, \alpha_{0M})$ . Other unobserved heterogeneity distributions are:

**Inverse Gaussian:**  $S(m; \mathbf{X}_m, \boldsymbol{\theta}) = \exp\left(\frac{1}{\sigma^2} \left[1 - (1 + 2\sigma^2 \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or})\right]^{\frac{1}{2}}\right)$

**Stable Distribution** (Hougaard, 1986):  $S(m; \mathbf{X}_m, \boldsymbol{\theta}) = \exp\left(\frac{1}{\sigma^2} \left[1 - (1 + c\sigma^2 \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or})\right]^{\frac{1}{c}}\right)$

**Mass-Point Distribution:**  $S(m; \mathbf{X}_m, \boldsymbol{\theta}) = \sum_{j=1}^J p_j \exp\left(-\xi_j \sum_{r=1}^m \exp(\mathbf{x}'_r \boldsymbol{\beta}) \alpha_{or}\right)$

where J equals the number of mass points and  $\sum_{j=1}^J p_j = 1$ . Rather than fixing the mean of the

mixing distribution to 1 for this distribution, empirical implementation is easier by instead fixing  $\theta_{01} = 1$ .

## 7. Multi-spell Discrete-time Models

Suppose that instead of a single duration we have multiple durations. For example, we may be interested in examining consecutive unemployment durations. The survivor function for the  $g^{\text{th}}$  spell satisfies

$$S^g(k_m) = P(T^g > k_m) = \prod_{r=1}^m P(T^g > k_r | T^g > k_{r-1})$$

We shall assume that

$$P(T^g > k_m^g | T^g > k_{m-1}^g) = \alpha_m^g(\mathbf{x}_m^g, \theta) \equiv \exp[-\exp(\mathbf{x}_m^{g'} \boldsymbol{\beta}^g) \alpha_{m0}^g]$$

or

$$P(T^g = k_m^g | T^g > k_{m-1}^g) = 1 - \alpha_m^g(\mathbf{x}_m^g, \theta) \equiv 1 - \exp[-\exp(\mathbf{x}_m^{g'} \boldsymbol{\beta}^g) \alpha_{m0}^g] .$$

Thus,

$$S^g(k_m^g) = P(T^g > k_m^g) = \prod_{r=1}^m \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g]$$

We will have data for up to G spells for an individual. If an individual completes the  $g^{\text{th}}$  spell with  $K^g = k^g$  that spell contributes

$$P(K^g = k_i^g) = \left\{ \prod_{r=1}^{k_i^g-1} \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g] \right\} \left( \exp[-\exp(\mathbf{x}_{k_i^g}^{g'} \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right)$$

to the likelihood function. If they are censored in the  $g^{\text{th}}$  spell at time  $k_i^g$  then that spell contributes

$$P(K^g = k_i^g) = \left\{ \prod_{r=1}^{k_i^g} \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g] \right\}$$

to the likelihood function. Let the spell indicator variables  $v_i^g = 1$  if an individual enters the  $g^{\text{th}}$

spell and zero otherwise,  $g=1, \dots, G$  and define the censor variables  $d_i^g = 1$  if the individual completes the  $g^{\text{th}}$  spell and zero otherwise,  $g=1, \dots, G$ . Then each individual contributes

$$L_i = \prod_{g=1}^G \left( \left\{ \prod_{r=1}^{k_i^g-1} \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g] \right\} \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^{g'} \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right)^{d_i^g} \right)^{v_i^g}$$

to the likelihood function which is

$$L = \prod_{i=1}^N L_i = \prod_{i=1}^N \prod_{g=1}^G \left[ \left\{ \prod_{r=1}^{k_i^g-1} \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g] \right\} \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^{g'} \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right)^{d_i^g} \right]^{v_i^g}$$

and the log-likelihood function equals

$$\begin{aligned} \ln(L) &= \sum_{i=1}^N \sum_{g=1}^G \left( \sum_{r=1}^{k_i^g-1} \ln \left\{ \exp[-\exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g] \right\} \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^{g'} \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right)^{d_i^g} \right)^{v_i^g} \\ &= \sum_{i=1}^N \sum_{g=1}^G v_i^g \left[ \left\{ - \sum_{r=1}^{k_i^g-1} \exp(\mathbf{x}_r^{g'} \boldsymbol{\beta}^g) \alpha_{0r}^g \right\} + d_i^g \ln \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^{g'} \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right) \right] \end{aligned}$$

Under the assumption that spell durations are independent, the likelihood decomposes and one can obtain estimates for this model by estimating  $G$  single spell models where all individuals with  $v_i^g = 1$  are included in the  $g^{\text{th}}$  estimation. If we instead assume that  $\boldsymbol{\beta}^g = \boldsymbol{\beta}$  and  $\alpha_{0m}^g = \alpha_{0m}$  for all  $g$  then we can “stack” observations for each spell within each individual and estimate a single spell duration model with log likelihood

$$\ln(L) = \sum_{i=1}^N \sum_{g=1}^G v_i^g \left[ \left\{ -\sum_{r=1}^{k_i^g-1} \exp(\mathbf{x}_r^g \boldsymbol{\beta}) \alpha_{0r} \right\} + d_i^g \ln \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^g \boldsymbol{\beta}) \alpha_{0k_i^g}] \right) \right]$$

An intermediate case would be a model that restricts  $\boldsymbol{\beta}^g = \boldsymbol{\beta}$  for all g but allows the baseline hazard parameters to be spell dependent.

$$\ln(L) = \sum_{i=1}^N \sum_{g=1}^G \sum_{r=1}^{k_i^g-1} \left\{ -v_i^g \exp(\mathbf{x}_r^g \boldsymbol{\beta}) \alpha_{0r}^g + d_i^g \ln \left( 1 - \exp[-\exp(\mathbf{x}_{k_i^g}^g \boldsymbol{\beta}) \alpha_{0k_i^g}^g] \right) \right\}$$

This case is similar to the continuous duration Cox regression model that stratifies the stacked data by spell. One could employ single spell discrete duration methods to estimate such a model by stacking the data and incorporating (G-1) ? M time-varying covariates that are of the form  $x_{gm} = I(K = k) \times I(S = g)$  where S is a variable denoting the particular spell.

Estimation becomes more complicated if we assume that for each duration

$P(T^g > k_m^g | T^g > k_m^g - 1)$  has the form

$$P(T^g > k_m^g | T^g > k_m^g - 1) = \alpha_m^g(\mathbf{x}_m^g, \boldsymbol{\theta}^g) \equiv \exp[-\xi^g \exp(\mathbf{x}_m^g \boldsymbol{\beta}^g) \alpha_{0m}^g]$$

where  $\xi^g$  are unobserved random variables which are assumed independent of the covariate processes  $\mathbf{x}_m^g$ ,  $g=1, \dots, G$ . In general, the  $\xi^g$  may be correlated with each other. Denote this joint distribution of the G x 1 vector  $\boldsymbol{\xi}$  by  $G(\boldsymbol{\xi}; \boldsymbol{\delta})$  where we have assumed that the distribution can be parameterized by the Q x 1 vector  $\boldsymbol{\delta}$ . The unconditional log likelihood function is obtained by integrating out  $\boldsymbol{\xi}$ . Thus we have

$$\ln(L) = \sum_{i=1}^N \int \ln \left( \prod_{g=1}^G \left[ \prod_{r=1}^{k_i^g-1} \left( \exp(-\xi^g \exp(\mathbf{x}_r^g \boldsymbol{\beta}^g) \alpha_{0r}^g) \right) \right] \left( 1 - \exp[-\xi^g \exp(\mathbf{x}_{k_i^g}^g \boldsymbol{\beta}^g) \alpha_{0k_i^g}^g] \right)^{d_i^g} \right]^{v_i^g} dG(\boldsymbol{\xi}; \boldsymbol{\delta}) \quad (31)$$

Estimates are obtained by maximizing () with respect to the  $\beta_g$ ,  $\alpha_{0mg}$ , and  $\delta$ . Maximum likelihood estimation may prove computationally intensive since the integral in (31) is typically multivariate. One may assume a mass-point specification for G where there are M types of individuals in the population and each type as a unique G x 1 vector  $\theta$  of location parameters.

Let  $p_q$  denote the proportion of the  $q^{\text{th}}$  type in the population,  $q=1, \dots, Q$ . Then the log-likelihood (31) becomes

$$\ln(L) = \sum_{i=1}^N \ln \left( \sum_{q=1}^Q p_q \prod_{g=1}^G \left[ \prod_{r=1}^{k_i^g-1} \left( \exp(-\xi_q^g \exp(\mathbf{x}_r^g \beta^g) \alpha_{0r}^g) \right) \right] \left( 1 - \exp[-\xi_q^g \exp(\mathbf{x}_{k_i^g}^g \beta^g) \alpha_{0k_i^g}^g] \right)^{d_i^g} \right]^{v_i^g} \right)$$

This likelihood is then maximized with respect to  $\beta^g$ ,  $\alpha_{0m}^g$ ,  $g=1, \dots, G$  and  $\xi_q$ , and  $p_q$   $q=1, \dots, Q$

where  $\sum_{q=1}^Q p_q = 1$ .

## **8. Competing Risk Models**

In many cases spells end for different reasons. Individuals may quit a job or be laid off, a person may die because of cancer or a heart attack, re-employment may occur into a job that is part-time or full-time. In such cases the explanatory variables may have differing effects on the relative probabilities of spells ending for particular reasons. Competing risks framework is meant to allow for this possibility. While without regressors it is not possible in general to distinguish models with correlated risks from those with independent risk, Heckman and Honore (1989) have derived sufficient conditions on the regressors which enable such identification. We

shall assume that such regressors exist. Also for simplicity in the discussion below we focus exclusively on the case of two risks. Extension to cases where the number of risks exceeds two is straightforward.

Competing risk models assume that there are two latent duration variables,  $T_1$  and  $T_2$ , which represent the time until the occurrence of the type 1 and type 2 events, respectively. What is observed, however, is only  $T = \min(T_1, T_2)$  and an indicator  $I\{T=T_1\}$ . This is referred to as the “identified minimum”. Thus, we know not only how long it took before at least one of the two types of events occurred but also which one it was. For example, if you were laid off from your job after  $T$  months of tenure, we know that you hadn’t quit and weren’t laid off before this time. We also know the reason for job spell ending (i.e. layoff). What we don’t know is when (if ever) you would have quit your job had you not been laid off first.

We assume that duration data are discrete and proceed by specifying the joint survivor function for the two latent durations  $T_1$  and  $T_2$  which is denoted by  $S(T_1, T_2)$ . In particular we assume that

$$P(T^1 > k_r^1 | T^1 > k_{r-1}^1) = \alpha_r^1(\mathbf{x}_r', \xi^1) \equiv \exp[-\xi^1 \exp(\mathbf{x}_r' \boldsymbol{\beta}^1) \alpha_{0r}^1] \quad (32)$$

and

$$P(T^2 > k_r^2 | T^2 > k_{r-1}^2) = \alpha_r^2(\mathbf{x}_r', \xi^2) \equiv \exp[-\xi^2 \exp(\mathbf{x}_r' \boldsymbol{\beta}^2) \alpha_{0r}^2] \quad (33)$$

where we assume that the variables  $\xi^1$  and  $\xi^2$  are unobserved and independent of the observed explanatory variables. Correlated risks arise in this model to the extent that  $\xi^1$  and  $\xi^2$  are

correlated. From (32) and (33) the latent survivor function satisfies

$$\begin{aligned} S(t_1, t_2 | \mathbf{X}, \xi^1, \xi^2) &= \prod_{t=1}^{t_1} \exp[-\xi^1 \exp(\mathbf{x}'_t \boldsymbol{\beta}^1) \alpha_{0t}^1] \prod_{t=1}^{t_2} \exp[-\xi^2 \exp(\mathbf{x}'_t \boldsymbol{\beta}^2) \alpha_{0t}^2] \\ &= \exp \left[ -\xi^1 \sum_t \exp(\mathbf{x}'_t \boldsymbol{\beta}^1) \alpha_{0t}^1 - \xi^2 \sum_t \exp(\mathbf{x}'_t \boldsymbol{\beta}^2) \alpha_{0t}^2 \right] \end{aligned}$$

where  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{\max(t_1, t_2)}\}$ . Let  $G$  be the distribution function for the unobservables  $\xi^1$  and  $\xi^2$ . Then the unconditional survivor function satisfies

$$S(k^1, k^2 | \mathbf{X}) = \int \exp \left[ -\xi^1 \sum_t \exp(\mathbf{x}'_t \boldsymbol{\beta}^1) \alpha_{0t}^1 - \xi^2 \sum_t \exp(\mathbf{x}'_t \boldsymbol{\beta}^2) \alpha_{0t}^2 \right] dG(\xi^1, \xi^2). \quad (34)$$

To construct the likelihood function in this case suppose for the  $i$ th individual the fail at time  $t$  due to cause 1. Then,

$$\begin{aligned} P(k = \min(k^1, k^2), I(k = k^1) = 1 | \mathbf{X}, \xi^1, \xi^2) &= \\ S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) &\left[ P(\{k-1 < T^1 \leq k\} \cap \{T^1 > T^2\} | \{T^1 > k-1\} \cap \{T^2 > k-1\}, \mathbf{X}, \xi^1, \xi^2) \right] = \\ &= S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) + \left[ P(k-1 < T^1 \leq k | \{T^1 > k-1\} \cap \{T^2 > k\}, \mathbf{X}, \xi^1, \xi^2) \right. \\ &+ P(\{k-1 < T^1 \leq k\} \cap \{T^1 > T^2\} | \{T^1 > k-1\} \cap \{k-1 < T^2 \leq k\}, \mathbf{X}, \xi^1, \xi^2) \left. \right] \\ &= (S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) - S(k-1, k | \mathbf{X}, \xi^1, \xi^2)) \\ &+ 1/2 \left[ S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) + S(k, k | \mathbf{X}, \xi^1, \xi^2) - S(k-1, k | \mathbf{X}, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}, \xi^1, \xi^2) \right] \\ &= (S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}, \xi^1, \xi^2)) + A(k | \mathbf{X}, \xi^1, \xi^2) \end{aligned}$$

where

$$A(k | \mathbf{X}, \xi^1, \xi^2) \equiv 1/2 \left[ S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) + S(k, k | \mathbf{X}, \xi^1, \xi^2) - S(k-1, k | \mathbf{X}, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}, \xi^1, \xi^2) \right].$$

In a similar manner we have

$$P(k = \min(k^1, k^2), I(k = k^2) = 1 | \mathbf{X}, \xi^1, \xi^2) = \left( S(k-1, k-1 | \mathbf{X}, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}, \xi^1, \xi^2) \right) + A(k | \mathbf{X}, \xi^1, \xi^2)$$

Let  $c_i^1$  ( $c_i^2$ ) be an indicator variable that equals 1 if the  $i$ th person spell ends for reason 1

(2) and let  $c_i^3$  be an indicator variable that equals one if the spell is right censored. Then the log

likelihood function for the competing risks model  $\text{Log}(\mathbf{L})$  satisfies:

$$\begin{aligned} \ln(\mathbf{L}) &= \sum_{i=1}^N c_i^1 \ln \left( \int \left\{ S(k-1, k-1 | \mathbf{X}_i, \xi^1, \xi^2) - S(k-1, k | \mathbf{X}_i, \xi^1, \xi^2) - A(k | \mathbf{X}_i, \xi^1, \xi^2) \right\} dG(\xi^1, \xi^2) \right) \\ &+ c_i^2 \ln \left( \int \left\{ S(k-1, k-1 | \mathbf{X}_i, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}_i, \xi^1, \xi^2) - A(k | \mathbf{X}_i, \xi^1, \xi^2) \right\} dG(\xi^1, \xi^2) \right) \\ &+ c_i^3 \ln \left( \int S(k-1, k-1 | \mathbf{X}_i, \xi^1, \xi^2) dG(\xi^1, \xi^2) \right) \end{aligned} \quad (35)$$

As we will discuss later, these methods were used by McCall (1996, 1997) to investigate the re-employment patterns of displaced workers in the United States. In particular, McCall (1996) used a competing risks framework to model joblessness durations that end due to re-employment into part-time or full-time jobs and how particular parameters of the US unemployment insurance (UI) system affect not only how long an individual remains unemployed but also the type (part-time versus full-time) of job that they are re-employed into.

In the United States, many unemployed individuals who qualify for UI benefits do not file a claim. McCall's (1996) analysis focused only on displaced workers who qualified for benefits. To allow for the possibility that changes in the parameters of the UI system will affect the choice to file a UI claim (see Anderson and Meyer, 1997 and McCall, 1995), McCall (1996)



modeled the claim filing choice and allowed for the possibility that unobservable determinants of that choice may be correlated with unobservable determinants (i.e.  $\xi^1$  and  $\xi^2$  in (35) ) of the re-employment rates into part-time and full-time jobs. Thus UI receipt was modeled by the dichotomous variable UI which equals 1 if an individual files a claim and 0 otherwise where  $\Pr(\text{UI}=1)$  has the functional form

$$P(\text{ui}=1)=1-\exp[-\xi'' \exp(\mathbf{z}'\boldsymbol{\delta})]$$

where  $\mathbf{z}$  is a vector of explanatory variables,  $\boldsymbol{\delta}$  is a vector of parameters and  $\xi''$  is an unmeasured variable that is uncorrelated with  $\mathbf{X}$  and  $\mathbf{z}$ . However,  $\xi''$  may be correlated with  $\xi^1$  and  $\xi^2$  in (35) and ui (along with its interaction with some variables in  $\mathbf{X}$ ) are added as explanatory variables in (35). Denote these variables by  $\mathbf{v}$ . The likelihood function for this model selectivity corrected competing risks model is

$$\begin{aligned} \ln(L) = & \sum_{i=1}^N c_i^1 \ln \left( \int P(\text{ui}_i=1|\mathbf{z}_i, \xi'')^{u_i} P(\text{ui}_i=0|\mathbf{z}_i, \xi'')^{1-u_i} \left\{ S(k-1, k-1 | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) - S(k-1, k | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) - A(k | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) \right\} dG(\xi^1, \xi^2, \xi'') \right) \\ & + c_i^2 \ln \left( \int P(\text{ui}_i=1|\mathbf{z}_i, \xi'')^{u_i} P(\text{ui}_i=0|\mathbf{z}_i, \xi'')^{1-u_i} \left\{ S(k-1, k-1 | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) - S(k, k-1 | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) - A(k | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) \right\} dG(\xi^1, \xi^2, \xi'') \right) \\ & + c_i^3 \ln \left( \int P(\text{ui}_i=1|\mathbf{z}_i, \xi'')^{u_i} P(\text{ui}_i=0|\mathbf{z}_i, \xi'')^{1-u_i} S(k-1, k-1 | \mathbf{X}_i, \text{ui}_i, \xi^1, \xi^2) dG(\xi^1, \xi^2, \xi'') \right) \end{aligned} \quad (36)$$

McCall(1996) used a mass-point specification for the unobserved heterogeneity distribution G which assumes that there are M types of individuals in the population with type m having the unique triplet  $(\xi_m^1, \xi_m^2, \xi_m'')$  of “location” points and composing  $p_m$  of the population,  $m=1, \dots, M$ , with

$$\sum_{m=1}^M p_m = 1. \text{ For this specification, the likelihood in (36) becomes}$$

$$\begin{aligned}
\ln(L) = & \sum_{i=1}^N c_i^1 \ln \left( \sum_{m=1}^M p_m P(u_i = 1 | \mathbf{z}_i, \xi_m^u)^{u_i} P(u_i = 0 | \mathbf{z}_i, \xi_m^u)^{1-u_i} \left\{ S(k-1, k | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) - S(k, k | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) - A(k | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) \right\} \right) \\
& + c_i^2 \ln \left( \sum_{m=1}^M p_m P(u_i = 1 | \mathbf{z}_i, \xi_m^u)^{u_i} P(u_i = 0 | \mathbf{z}_i, \xi_m^u)^{1-u_i} \left\{ S(k, k-1 | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) - S(k, k | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) - A(k | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) \right\} \right) \\
& + c_i^3 \ln \left( \sum_{m=1}^M p_m P(u_i = 1 | \mathbf{z}_i, \xi_m^u)^{u_i} P(u_i = 0 | \mathbf{z}_i, \xi_m^u)^{1-u_i} S(k-1, k-1 | \mathbf{X}_i, u_i, \xi_m^1, \xi_m^2) \right) \\
(6)
\end{aligned}$$

Two particular unemployment insurance parameters of interest to McCall(1996) were the weekly benefit amount and the disregard amount which we discuss further in chapter XX.

## 9. General Discrete-Time Life History Models

More generally, we consider a discrete-time life history process that is characterized by a discrete-valued state space  $S$  and the following conditional transition probabilities  $P^r(F_{k-1})$  for  $r \in S$  which represents the conditional probability that the process is in state  $r$  at time  $k$  given the history of the process (information) up to (discrete) time  $k-1$ . The history would include not only the past history of transitions of the process but also the history of, possibly time-varying, covariates. If a more coarse history is observed,  $G_k$ , with  $G_k \supset F_k$  we will instead have

$E_{G_{k-1}}(P^r(F_{k-1}))$ . For example, a particular covariate that affects the transition probabilities may not be observed. A specific form of this probability may be

$$P^r(F_{k-1}) = 1 - \exp \left( -\xi_{s(k-1)}^{rg^s(k-1)} \exp(\mathbf{x}_k' \boldsymbol{\beta}_{s(k-1)}^{g^s(k-1)}) \alpha_{0m^s(k-1)}^{g^s(k-1)} \right)$$

where  $s(k-1)$  is the state occupied at time  $k-1$ ,  $g^s(k-1)$  represents how many visits to state  $s$  have occurred by  $k-1$  with

$$k = \sum_{s=1}^r \sum_{g=1}^{g(s)} k_s^g$$

and  $k_s^g$  represents the time spent in state  $s$  during its  $g^{\text{th}}$  visit. Note that the  $s$ ,  $g(s)$  and  $k_s^g$  are random variables at time 0. Now,

$$\sum_{s=1}^S P^r(F_{k-1}) = \sum_{s=1}^S 1 - \exp\left(-\xi_s^{rg(s)} \exp(\mathbf{x}_k \boldsymbol{\beta}_s^{g(s)}) \alpha_{s0k_s^g}^{g(s)}\right) = 1$$

and by convention we fix the probability of remaining in the state at  $t-1$  to

$$P^{s(k-1)}(F_{k-1}) = 1 - \sum_{j \neq s(t-1)} \left[ 1 - \exp\left(-\xi_j^{rg(j)} \exp(\mathbf{x}_k' \boldsymbol{\beta}_j^{g(j)}) \alpha_{j0k_j^g}^{g(j)}\right) \right].$$

Suppose that we observe only  $G_{k-1}$ , then

$$P^r(G_{k-1}) = \int \left\{ 1 - \exp\left(-\xi_s^{rg(s)} \exp(\mathbf{x}_k' \boldsymbol{\beta}_s^{g(s)}) \alpha_{s0k_s^g}^{g(s)}\right) \right\} d\mathcal{P}_{G_{k-1}}.$$

In particular we assume that durations in each state are observed as well as all  $\mathbf{x}$ 's. Only the variables  $\xi_s^{rg(s)}$  are unobserved and the  $\xi_s^{rg(s)}$  are  $F_0$  measurable and distributed independently of the  $\mathbf{x}$ 's. Then

$$P^r(G_{k-1}) = \int \left( P^r(F_{k-1}) L(F_{k-1} | F_0) \right) dB(\boldsymbol{\xi})$$

where  $L(F_{k-1} | F_0)$  represents the probability (“likelihood”) of observing a particular history  $F_{k-1}$  given  $F_0$  and  $B(\boldsymbol{\xi})$  is the distribution of the vector  $\boldsymbol{\xi}$ . Essentially we “integrate out” the vector of

variables  $\xi$ . Suppose for individual  $i$  we observe the sequence of times spent in states and the  $\mathbf{x}$ 's.

Further suppose that the distribution  $B$  can be characterized by the parameter  $\delta$ . Since we have

$$L(F_{k-1} | F_0) = \prod_{j=1}^{k-1} L(F_j | F_{j-1}) =$$

Then the individual contribution to the likelihood function equals

$$\begin{aligned} L_i &= \int \left( \prod_{j=1}^{k_i} L(F_j | F_{j-1}) \right) dB(\xi, \delta) \\ &= \int \left( \prod_{j=1}^{k_i} P^{r(j)}(F_{j-1}) \right) dB(\xi, \delta) \\ &= \int \prod_{j=1}^{k_i} \left[ 1 - \exp \left( -\xi_{s(j-1)}^{rg(s(j-1))} \exp(\mathbf{x}_j' \boldsymbol{\beta}_{s(j-1)}^{g(s(j-1))}) \alpha_{s(j-1)0k_{j-1}^s}^{g(s(j-1))} \right) \right]^{I(s(j) \neq s(j-1))} \\ &\quad \left[ 1 - \sum_{r \neq s(j-1)}^S \left[ 1 - \exp \left( -\xi_{s(j-1)}^{rg(s(j-1))} \exp(\mathbf{x}_j' \boldsymbol{\beta}_{s(j-1)}^{g(s(j-1))}) \alpha_{s(j-1)0k_{j-1}^s}^{g(s(j-1))} \right) \right] \right]^{I(s(j) = s(j-1))} dB(\xi, \delta) \end{aligned}$$

The log-likelihood function is then

$$\ln(L) = \sum_{i=1}^N \ln(L_i)$$

which is maximized with respect to the parameters  $\boldsymbol{\beta}_s^g$ ,  $\alpha_{s0m}^g$ , and  $\delta$ .

One example which applies these discrete-time life history methods would be the analysis of employment to unemployment and unemployment to employment transitions when workers may have access to unemployment insurance. In the United States and Canada, workers can still receive benefits while working for jobs with low earnings. In particular, in Canada workers are allowed to earn the maximum of \$50 or 25% of their weekly benefit amount with no reduction in

unemployment insurance benefits. For any additional earnings above this amount, weekly benefits are reduced dollar for dollar. Given the possibility that benefits may be received while employed or unemployed, it may be desirable to specify an econometric model with four distinct states: employed-benefits (EB), employed-no benefits (ENB), unemployed-benefits (UB), unemployed-no benefits (UNB).

The transitions UNB $\rightarrow$ ENB , UB $\rightarrow$ ENB, ENB $\rightarrow$ UNB and UNB $\rightarrow$ UB would commonly be observed where the latter transition occurs when an already unemployed individual files for a claim or when a newly unemployed individual satisfies a waiting period that is required by law (e.g. in Canada there is a two week waiting period before benefits can be received). Other transitions that are possible but perhaps less common are the UB $\rightarrow$ UNB transition that occurs when an individual exhausts unemployment insurance benefits or is disqualified from receiving benefits for some reason (e.g. inadequate job search). The UB $\rightarrow$ EB occurs when an individual receiving unemployment insurance benefits begins a low-earnings job that allows them to continue receiving benefits. The transitions EB $\rightarrow$ UB and ENB $\rightarrow$ UB occur when an individual who is in the benefit period from a past job loss and has not exhausted all benefits loses their current job. The EB $\rightarrow$ ENB may occur either when a person who is working and receiving benefits exhausts their benefits or when a person who is working and receiving benefits experiences a sufficiently large increase in earnings that disqualifies them from receiving further benefits while working.

Suppose we have panel data that follows a sample of individuals from the time they lose a job forward. Thus, all individuals begin in state UNB. Here, our goal is simply to demonstrate how a likelihood function would be constructed from the individual life histories. So, consider a

particular individual (i) who files for and receives benefits after satisfying the two week waiting period, receives benefits without working for then five weeks before working part-time. The individual still receives benefits while working and this job last four weeks. The individual then is unemployed (and receiving UI benefits) for eight weeks before finding another part-time job. The individual still receives benefits while working at this new part-time job and this job last eight weeks before the individual is made a (permanent) full-time job offer which precludes receiving further benefits. This particular history is portrayed in Figure 1.

The components of the contribution to the likelihood for individual i would be

$$P_{UNB}^{UNB}(1) = 1 - \sum_{r \neq \{UNB\}}^S \left[ 1 - \exp\left(-\xi_{UNB}^{r1} \exp(\mathbf{x}_1' \boldsymbol{\beta}_{UNB}^{r1}) \alpha_{UNB01}^{r1}\right) \right]$$

$$P_{UNB}^{UB}(2) = 1 - \exp\left(-\xi_{UNB}^{UB1} \exp(\mathbf{x}_2' \boldsymbol{\beta}_{UNB}^{UB1}) \alpha_{UNB02}^{UB1}\right)$$

for the unemployment period until benefits are received,

$$P_{UB}^{UB}(3) = 1 - \sum_{r \neq \{UB\}}^S \left[ 1 - \exp\left(-\xi_{UB}^{r1} \exp(\mathbf{x}_3' \boldsymbol{\beta}_{UB}^{r1}) \alpha_{UB01}^{r1}\right) \right]$$

⋮

$$P_{UB}^{UB}(5) = 1 - \sum_{r \neq \{UB\}}^S \left[ 1 - \exp\left(-\xi_{UB}^{r1} \exp(\mathbf{x}_5' \boldsymbol{\beta}_{UB}^{r1}) \alpha_{UB03}^{r1}\right) \right]$$

$$P_{UB}^{EB}(6) = 1 - \exp\left(-\xi_{UB}^{EB1} \exp(\mathbf{x}_6' \boldsymbol{\beta}_{UB}^{EB1}) \alpha_{UB04}^{EB1}\right)$$

for the benefit receipt period until the first part-time job is received,

$$\begin{aligned}
P_{EB}^{EB}(7) &= 1 - \sum_{r \neq \{EB\}}^S \left[ 1 - \exp\left(-\xi_{EB}^{r1} \exp(\mathbf{x}_7' \boldsymbol{\beta}_{EB}^{r1}) \alpha_{EB01}^{r1}\right) \right] \\
&\vdots \\
P_{EB}^{EB}(9) &= 1 - \sum_{r \neq \{EB\}}^S \left[ 1 - \exp\left(-\xi_{EB}^{r1} \exp(\mathbf{x}_9' \boldsymbol{\beta}_{EB}^{r1}) \alpha_{EB03}^{r1}\right) \right] \\
P_{EB}^{UB}(10) &= 1 - \exp\left(-\xi_{EB}^{UB1} \exp(\mathbf{x}_{10}' \boldsymbol{\beta}_{EB}^{UB1}) \alpha_{EB04}^{UB1}\right)
\end{aligned}$$

for the first period of working on claim,

$$\begin{aligned}
P_{UB}^{UB}(11) &= 1 - \sum_{r \neq \{UB\}}^S \left[ 1 - \exp\left(-\xi_{UB}^{r2} \exp(\mathbf{x}_{11}' \boldsymbol{\beta}_{UB}^{r2}) \alpha_{UB01}^{r2}\right) \right] \\
&\vdots \\
P_{UB}^{UB}(17) &= 1 - \sum_{r \neq \{UB\}}^S \left[ 1 - \exp\left(-\xi_{UB}^{r2} \exp(\mathbf{x}_{17}' \boldsymbol{\beta}_{UB}^{r2}) \alpha_{UB07}^{r2}\right) \right] \\
P_{UB}^{EB}(18) &= 1 - \exp\left(-\xi_{UB}^{EB2} \exp(\mathbf{x}_{18}' \boldsymbol{\beta}_{UB}^{EB2}) \alpha_{UB08}^{EB2}\right)
\end{aligned}$$

for the second period of unemployed benefit receipt,

$$\begin{aligned}
P_{EB}^{EB}(19) &= 1 - \sum_{r \neq \{EB\}}^S \left[ 1 - \exp\left(-\xi_{EB}^{r2} \exp(\mathbf{x}_{19}' \boldsymbol{\beta}_{EB}^{r2}) \alpha_{EB01}^{r2}\right) \right] \\
&\vdots \\
P_{EB}^{EB}(21) &= 1 - \sum_{r \neq \{EB\}}^S \left[ 1 - \exp\left(-\xi_{EB}^{r2} \exp(\mathbf{x}_{21}' \boldsymbol{\beta}_{EB}^{r2}) \alpha_{EB03}^{r2}\right) \right] \\
P_{EB}^{ENB}(22) &= 1 - \exp\left(-\xi_{EB}^{ENB2} \exp(\mathbf{x}_{22}' \boldsymbol{\beta}_{EB}^{ENB2}) \alpha_{EB04}^{EN2}\right)
\end{aligned}$$

for the second period of working on claim, and

$$\begin{aligned}
P_{ENB}^{ENB}(23) &= 1 - \sum_{r \neq \{ENB\}}^S \left[ 1 - \exp\left(-\theta_{ENB}^{r1} \exp(\mathbf{x}(22)' \boldsymbol{\beta}_{ENB}^{r1}) \alpha_{ENB01}^{r1}\right) \right] \\
&\vdots \\
P_{ENB}^{ENB}(23+K) &= 1 - \sum_{r \neq \{ENB\}}^S \left[ 1 - \exp\left(-\theta_{ENB}^{r1} \exp(\mathbf{x}(22+K)' \boldsymbol{\beta}_{ENB}^{r1}) \alpha_{ENB0K}^{r1}\right) \right]
\end{aligned}$$

for the final period of working “off claim” where we assume that the individual is right censored after K period on the job.

Thus, the contribution to the likelihood equals of this individual

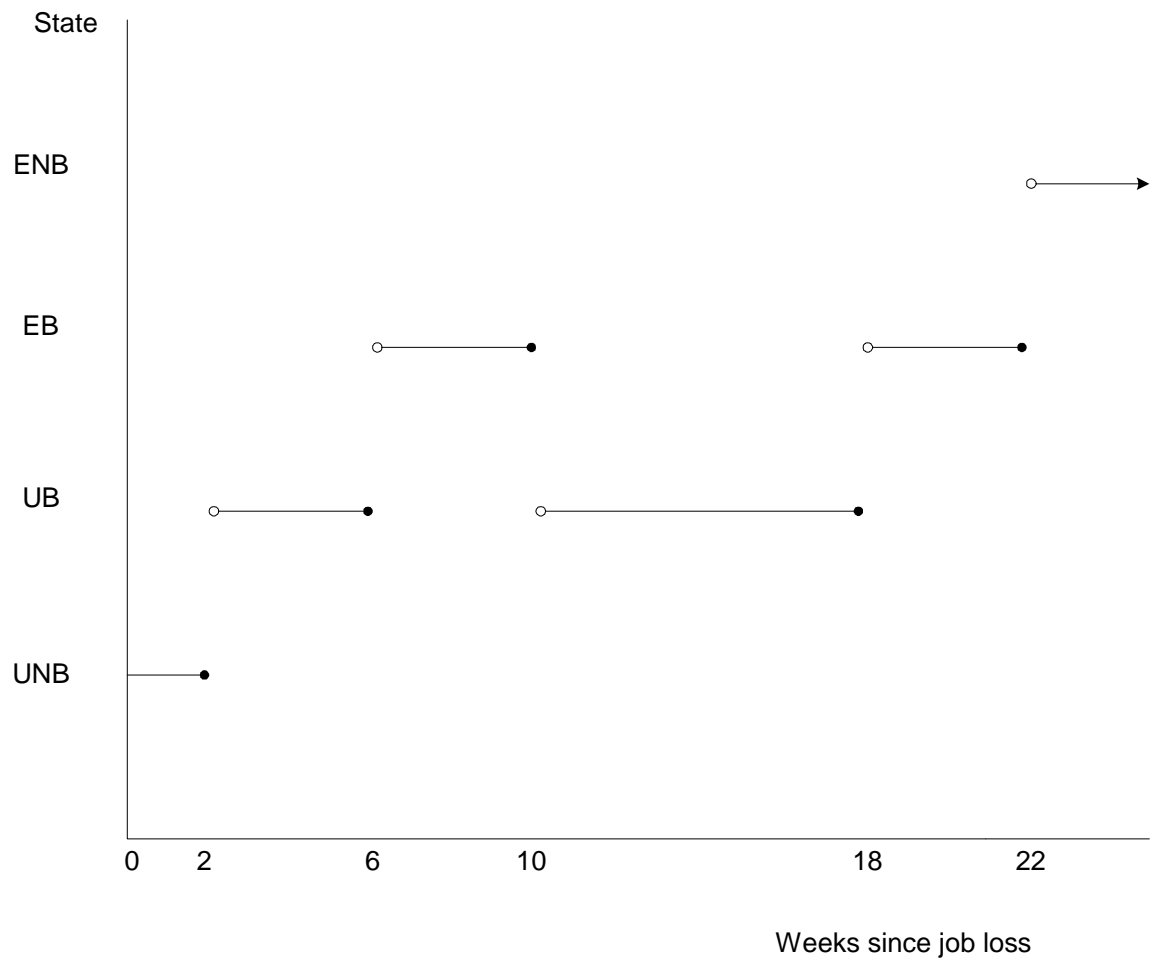
$$\begin{aligned}
L_i &= \int \left( P_{UNB}^{UNB}(1) P_{UNB}^{UB}(2) \left\{ \prod_{k=3}^5 P_{UB}^{UB}(k) \right\} P_{UB}^{EB}(6) \left\{ \prod_{k=7}^9 P_{EB}^{EB}(k) \right\} P_{EB}^{UB}(10) \right. \\
&\quad \left. \left\{ \prod_{k=11}^{17} P_{UB}^{UB}(k) \right\} P_{UB}^{EB}(18) \left\{ \prod_{k=19}^{21} P_{EB}^{EB}(k) \right\} P_{EB}^{ENP}(22) \left\{ \prod_{k=23}^{22+K} P_{ENB}^{ENB}(k) \right\} \right) dB(\boldsymbol{\xi}, \boldsymbol{\delta})
\end{aligned}$$

In this setup we have allowed for the baseline hazards parameters and coefficients associated with the explanatory variables to depend on the total number of previous visits to that state. More general forms of state dependence could also be incorporated into the model. For example, the transition probabilities may depend not only on the number of times the state was previously visited but also on the time spent in the state on each previous visits.



**Figure 1**

An Example of a Sample Path



## **10. Specification Tests for Duration Models**

To conclude this chapter we briefly consider specification testing for duration models. While there are several methods to potentially test the validity of a model specification in hazard models, one particularly useful test is based on the notion that if the model is correctly specified then

$$M_i(t) = \int_0^t Y_i(s)Z_i(s)dN_i(s) - \int_0^t Y_i(s)Z_i(s)\lambda_0(s)\exp(\mathbf{x}_i'\boldsymbol{\beta})ds$$

is a martingale.<sup>2</sup> Thus,

$$\hat{M}_i(t) = \int_0^t Y_i(s)Z_i(s)dN_i(s) - \int_0^t Y_i(s)Z_i(s)\hat{\lambda}_0(s)\exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})ds$$

which is based on the sample estimates of  $\lambda_0(s)$  and  $\boldsymbol{\beta}$  should be approximately a martingale if the model is correctly specified. Using the fact that the estimates are consistent and appealing to the martingale central limit theorem, one can construct Chi-square tests of model specification. Moreover, graphical assessment is feasible since under the null hypothesis that the model is correctly specified  $\hat{M}_i(t)$  is “approximately” a martingale for all  $i$  and  $t$  and, hence, plots of weighted sums of  $\hat{M}_i(t)$  should appear as “white-noise” or patternless (See Arjas, 1989). The remainder of this section follows McCall (1994) and focuses on discrete-time duration models. In the discrete case we work with martingale difference sequences. Let  $N_i(k) = I\{K_i=k, C_i>k\}$  where  $C_i$  denotes a censoring-time variable. Then,

$$x_i(k) = N_i(k) - I\{K > k-1\}I\{C > k\}[1 - \alpha_k(\mathbf{x}_k, \theta)]$$

forms a martingale difference sequence. These martingale differences can be standardized by the

---

<sup>2</sup> For additional specification diagnostic methods for duration models see Schoenfeld (1980), Wei (1984), and Lancaster (1990).

stochastic variance process  $v_i(k)$  where

$$v_i(k) = I\{K > k-1\} I\{C > k\} \alpha_k(\mathbf{x}_k, \theta) [1 - \alpha_k(\mathbf{x}_k, \theta)].$$

Suppose the data is observed over M periods and let  $\alpha$  denote the vector of  $\alpha_k(\mathbf{x}_k, \theta)$  for  $k=1, \dots, M$ . If the model is estimated by maximum likelihood then under suitable conditions (See McCall, 1994) the test statistic

$$\chi^2 = N^{-1} \hat{\mathbf{x}}' \hat{\Sigma}_0 \hat{\mathbf{x}}$$

is asymptotically chi-square distributed with  $\text{rank}(\hat{\Sigma}_0)$  degrees of freedom where

$$\hat{\mathbf{x}} = \sum_{i=1}^N \hat{\mathbf{x}}_i \text{ and } \hat{\mathbf{x}}_i \text{ are M vectors of } x_i(k), k=1, \dots, M \text{ evaluated at } \hat{\theta} \text{ and}$$

$$\Sigma_0 = \mathbf{V}_0 - E_p(\partial \mathbf{p}_0 / \partial \theta') E_p(\mathbf{I}_0)^{-1} E_p(\mathbf{s}_0 \mathbf{x}') - \left\{ E_p(\partial \mathbf{p}_0 / \partial \theta') E_p(\mathbf{I}_0)^{-1} E_p(\mathbf{s}_0 \mathbf{x}') \right\}' + E_p(\partial \mathbf{p}_0 / \partial \theta') E_p(\mathbf{I}_0)^{-1} E_p(\partial \mathbf{p}_0 / \partial \theta)'$$

with  $\mathbf{s}_0$  the score vector of the log-likelihood function,  $\mathbf{I}_0$  the Hessian matrix of second derivatives of the log-likelihood function and  $\mathbf{V}_0$  equal to the M x M diagonal matrix with  $(m,m)^{\text{th}}$  element

$$E_p v_i(m) = E_p \left( I\{K > m-1\} I\{C > m\} \alpha_m(\mathbf{x}_m, \theta) [1 - \alpha_m(\mathbf{x}_m, \theta)] \right).$$

## References

- Andersen, P. and O. Borgan. (1985). "Counting Process Models for Life History Data: A Review." Scandinavian Journal of Statistics, 12, 97-158.
- Anderson, P., Borgan, O., Gill, R. and N. Keiding. (1993). Statistical Models Based on Counting Processes, (Springer-Verlag: New York, NY).
- Arjas, E. (1988). "A Graphical Model for Assessing Goodness of Fit in Cox's Proportional Hazards Model." Journal of the American Statistical Association, 83, 204-212.
- Arjas, E. (1989) "Survival Models and Martingale Dynamics.", Scandinavian Journal of Statistics, 16, 177-225.
- Bremaud, P. (1981), Point Processes and Queues: Martingale Dynamics, (Springer-Verlag, New York, NY).
- Breslow, N. (1974). "Covariance Analysis of Censored Survival Data." Biometrics, 30, 89-99.
- Cox, D. (1972). "Regression Models with Life Tables (with discussion)." Journal of the Royal Statistical Society B, 34, 187-220.
- Cox, D., (1975). "Partial Likelihood." Biometrika, 62, 269-276.
- Devine, T. and N. Kiefer (1991). Empirical Labor Economics, (Oxford University Press: New York).
- Efron, B. (1977). "The Efficiency of Cox's Likelihood Function for Censored Data." Journal of the American Statistical Association, 72, 557-565.
- Fleming, T. and D. Harrington (1991), Counting Processes and Survival Analysis, (John Wiley and Sons: New York).
- Han, A. and J. A. Hausman (1990). "Flexible Parametric Estimation of Duration and Competing

- Risk Models.” Journal of Applied Econometrics **5**(1): 1-28.
- Heckman, J. and B. Honoré (1989). "The Identifiability of the Competing Risks Model." Biometrika **76**, 325-330.
- Heckman, J. and B. Singer( 1984). “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data.” Econometrica, **52**, 271-320.
- Hougaard, P. (1986). “Survival Models for Heterogeneous Populations Derived from Stable Distributions.” Biometrika, **73**, 387-396.
- Kiefer N. (1988). "Economic Duration Data and Hazard Functions." Journal of Economic Literature, **26**. 646-679.
- Klein, J. and M. Moeschberger (1997) Survival Analysis: Techniques for Censored and Truncated Data, (Springer: New York, NY).
- Lancaster, T. (1979). “Econometric Models for the Duration of Unemployment." Econometrica, **47**, 939-956.
- Lancaster, T. (1990). The Econometric Analysis of Transition Data, (Cambridge University Press: Cambridge, United Kingdom).
- McCall, B. (1994). “Specification Models for Duration Models.” Journal of Econometrics, **60**, 293-312.
- McCall, B. (1995). “The Impact of Unemployment Insurance Benefit Levels on Reciprocity.” Journal of Business and Economic Statistics, **13**, 189-198.
- McCall, B. (1996). “Unemployment Insurance Rules, Joblessness, and Part-Time Work.” Econometrica **64**: 647-682.

- McCall, B. (1997). "The Determinants of Full-time versus Part-time Re-employment Following Job Displacement." Journal of Labor Economics, 15, 714-734.
- Meyer, B. (1990). "Unemployment Insurance and Unemployment Spells." Econometrica, 58:757-782.
- Prentice, R. and L. Gloeckler (1978). "Regression Analysis of Grouped Survival Data with Applications to Breast Cancer Data." Biometrics, 34, 57-67.
- Schoenfeld, D. (1980). "Chi-squared Goodness-of-fit Tests for the Proportional Hazards Model." Biometrika, 67, 145-153.
- Wei, L. (1984). "Testing the Goodness of Fit for Proportional Hazards Model with Censored Observations." Journal of the American Statistical Association, 79, 649-652.